WILLKIE FARR & GALLAGHER

1875 K Street, NW
Washington, DC 20006

Tel: 202 303 1000
Fax: 202 303 2000

May 23, 2003

Ms. Marlene H. Dortch
Secretary
Federal Communications Commission
445 12th Street, S.W.
Washington, DC 20554

Re: *Ex Parte* Presentation in CC Docket Nos. 99-273, 96-98 and 96-115

Dear Ms. Dortch:

On May 22, 2003, representatives of the Association of Directory Publishers ("ADP") met with Marcy Greene, William Kehoe, Daniel Shiman, and Ann Stevens of the Wireline Competition Bureau to discuss the above-referenced proceedings. ADP urged the Commission to deny the outstanding petitions for reconsideration. Copies of the attached document were distributed at the meeting. ADP was represented by Theodore Whitehouse and the undersigned. ADP President R. Lawrence Angove and ADP member Joseph Walsh also attended the meeting.

Pursuant to Section 1.1206(b)(2) of the Commission's rules, 47 C.F.R. § 1.1206(b)(2), this letter is being filed electronically for inclusion in the public record of the above-referenced proceedings.

Sincerely,

/s/
Jennifer K. Ashworth

Attachment

cc: Marcy Greene
William Kehoe
Daniel Shiman
Ann Stevens

Competition Between Networks: A Study of the Market for Yellow Pages

Boston University Industry Studies Project Working Paper #104

Marc Rysman*
Boston University

September 9, 2002

Abstract

This paper estimates the importance of network effects in the market for Yellow Pages. I estimate three simultaneous equations: consumer demand for usage of a directory, advertiser demand for advertising and a publisher's first-order condition (derived from profit-maximizing behavior). Estimation shows that advertisers value consumer usage and that consumers value advertising, implying a network effect. I find that internalizing network effects would significantly increase surplus. As an application, I consider whether the market benefits from monopoly (which takes advantage of network effects) or oligopoly (which reduces market power). I find that a more competitive market is preferable.
JEL L1, L4. Keywords: Network Effects, Discrete Choice Demand, Antitrust.

1 Introduction

This paper measures the importance of network effects in a particular market, the market for Yellow Pages directories. Yellow Pages are a network good because the value of the product depends (indirectly) on how many consumers use the product. This interdependency occurs because consumers value directories based on how much information and advertising are in a

*270 Bay State Rd., Boston, MA 02215, mrysman@bu.edu. I thank Peter Arcidiacono, Ray Deneckere, John Kennan, Sam Kortum, Kevin Lang, Aviv Nevo, Larry Samuelson, numerous seminar groups and especially Phil Haile for advice and encouragement. Also, the comments of three anonymous referees and the editor significantly improved this paper. This research would not be possible without the generosity of National Yellow Pages Monitor, Claritas, Inc., and the Yellow Pages Publishers Association. The Christensen Award in Empirical Economics provided important financial support.

directory. Meanwhile, retailers place more advertising in a directory if there are more consumers of a directory. Together, consumer behavior and advertiser behavior create a network effect. In fact, telephone company directories tend to have much higher prices, larger books and more usage then independent producers, suggesting that network effects are important in determining market structure. [1] Because data is available on consumer usage as well as on prices and quantities of advertising, data is available on "both sides" of the feedback loop. This feature allows for the explicit estimation of a feedback loop in a way that has not been done before.

To accomplish this goal, I estimate two demand curves simultaneously: The first is a consumer demand curve for directory usage as a function of advertising. The second is an advertiser demand curve for advertising as a function of consumer usage. I find that the amount that consumers use a directory increases in the directory's level of advertising. I also find that retailer demand for advertising in a directory increases in the amount that consumers use the directory. These two results imply that a network effect exists. In order to calculate equilibrium outcomes, I estimate a first-order condition derived from profit-maximizing behavior by the publisher. I measure the importance of the network effect by looking at how much potential surplus is forgone due to the market's failure to fully account for the network effect. I find that the amount is large relative to the amount of deadweight loss resulting from imperfect competition and relative to the amount of surplus realized in the market equilibrium.

To further explore the model, I consider the welfare tradeoff between competition and monopoly under network effects. Network effects imply that there is a welfare gain to coordinating economic activity on the same standard. However, if that standard is proprietary, the owner can wield significant market power. In that case, it is an empirical question whether welfare is maximized under competition or standardization. Using estimates from the structural model, I calculate equilibrium outcomes for different numbers of competitors and test if welfare

[1]A common way for people who work in the Yellow Pages industry to convey the profitability of their product is to compare it to the profitability of illicit narcotics. In interviews with the author, one person said "We earn more money than anyone this side of the Cali drug cartel." Another said "We like to say that we are the second most profitable industry in the world." Statistical evidence of profitability is presented in Section 3.1.

increases. I find that network effects are not so strong as to counteract the benefits of entry. The results show that the entry of independent publishers improves welfare, so rules that force telephone company publishers to facilitate entry are welfare-enhancing. [2]

The methodology developed here is also interesting because it could be applied to other markets. The methodology is relevant to any industry characterized by indirect network effects and incompatible networks. A particularly topical example is the market for operating systems. Like Yellow Pages directories, computer operating systems exhibit indirect network effects in the sense that higher consumer usage of a particular operating system leads to more available software and vice versa. Also, directories and operating systems are both incompatible networks in the sense that an advertisement placed in one directory confers no benefit on the user of another directory, just as software designed for one operating system confers no benefit on the user of a different operating system.

2 Related Literature

Network effects and positive feedback loops are the subject of increasing attention in both the academic and popular press. The theoretical literature on network effects begins with Katz and Shapiro (1985), who introduce the concept of indirect network effects under the name *hardware/software paradigm.* Hardware becomes more valuable when more compatible software is supplied, and the amount of software available depends on the amount of hardware

[2] This exercise is relevant for policy because the decision about whether or not to open the market to competition is now in question. Although independent publishers have existed in the United States almost since the beginning of the industry, telephone company publishers have an advantage because they have better access to customer data. However, a recent U.S. Supreme Court ruling established that White Pages are not copyrightable, effectively opening the market to competition (*Feist Publications, Inc. v. Rural Tel. Service Co.499 U.S. 340, 1991).* This outcome seems to have been widely expected and business practice adjusted well in advance of the ruling. Also, the Telecom Act of 1996 includes a provision that telephone companies make their listings available at "reasonable rates". Similar policies have recently been enacted in Europe and Australia. See http://www.accc.gov.au/media/mr1997/telstra.htm

that consumers purchase. In the case of Yellow Pages, the "hardware" is the directory and the "software" is the advertising. The first formal models of indirect network effects appear in Chou and Shy (1990) and Church and Gandal (1992). Chou and Shy (1990) and Economides and Flyer (1997) study the welfare implications of entry and find that the effects of entry depend on parameter values. Surveys of the literature on network effects are Katz and Shapiro (1994) and Economides (1996).

A number of recent theoretical papers study the general topic of two-sided markets. Rochet and Tirole (2002) provide a widely applicable model and discuss markets for advertising, credit cards, software and web portal usage. Anderson and Coate (2001) and Stegeman (2002) study broadcast markets in which retailers pay for advertising to reach consumers, where typically consumers dislike advertising. Armstrong (2002) surveys this literature and provides a synthesis. Armstrong shows for Yellow Pages that it is straightforward to write a model in which directories are distributed for free and retailers pay for advertising because each consumer generates more advertising revenue than the cost of manufacturing a directory.

The empirical literature on network effects can be traced back to Gandal (1994), Greenstein (1993) and Saloner and Shepard (1995). These papers evaluated the importance of installed base or network size, typically in reduced-form equations. Very little work studies positive feedback loops. For instance, Ohashi (2000) and Park (2000) study the dynamics of network effects in the VCR market but abstract from the video rental market, presumably because of data constraints. Also, most work on network effects has focused on high-technology products where, as Goolsbee and Klenow (1999) make clear with the case of personal computers, it is often difficult to distinguish between network effects and the effect of consumers learning from each other.

A paper that does explicitly estimate a positive feedback loop is Gandal, Kende and Rob (2000). They study the entry decisions of producers of compact disk players and producers of compact disks. A major difference from my work is that they study firms producing for a compatible standard. Because the standard is non-proprietary, there are no benefits to having a small number of firms so the application of the model studied here does not arise.

Two empirical models that are similar to the one here are by Rosse (1970) and Berry and Waldfogel (1999). Although Rosse's model is designed to identify the cost curve of a newspaper as opposed to measure network effects, his model does allow him to measure the feedback between readership and advertising. However, one would expect readers' valuation of newspaper advertisements to be ambiguous. [3] One reason that I choose to focus on Yellow Pages directories are that they are valuable explicitly because of the advertisements. Berry and Waldfogel analyze the effects of entry by radio stations. They also estimate consumer demand for radio play and retailer demand for advertising, although they lack some station specific data (e.g. listenership, price, geographic coverage) and so are constrained from addressing some of the issues explored here.

3 Industry and Data Characteristics

As stated in the introduction, the Yellow Pages industry is an excellent one for study both because of the structure of the market and because of the availability of excellent data. Section 3.1 discusses industry characteristics and how they motivate important assumptions. Section 3.2 reviews the data set.

3.1 Industry Characteristics

The Yellow Pages generated $11.5 billion in sales in 1997 (Elliott, 1998). Yellow Pages directories published by telephone companies seem to earn exceptional profits. A directory serving 240,000 people averages over $6,000 per page in revenue from display advertisements alone. The average size of such a book is 621 pages so the book brings in around $3.8 million in revenue from display advertisements. Industry sources estimate that variable costs of production

[3] In Rosse's paper, the parameter that captures the effect of advertising on readership is estimated to be positive but not significant.

for a book of this size would be less than $1 million. Indeed, Yellow Pages industry sources estimate that profits represent 35-45% of revenue. In the break-up of AT&T, the court assigned Yellow Pages publishers to local phone companies in order to hold down local rates. Public service commissions estimate that if local phone companies did not receive Yellow Pages profits, local rates would have to increase between $1.80 and $3.50 per line per month. (NARUC, 1994-1995).

While the Yellow Pages industry has traditionally been dominated by telephone companies that never enter each other's markets, there is some competition from independent publishers. Independent publishers print 38% of the directories in this data set. Figure 4 shows that almost 60% of the population in this data set receives 2 or 3 directories. This number is not just due to publishers that distribute overlapping directories: Figure 4 shows that more than 50% receive directories from more than 1 publisher. That is, the median person receives two directories from separate publishers. However, independent publishers have not been overly successful. Directories associated with telephone companies average 6.42 references per household per month, while the same number at independent directories is only 1.32. Directories associated with telephone companies are on average almost twice the size (746 pages compared to 413) and charge about twice as much ($2,014 to $1,221 for a double-quarter column advertisement) as independent publishers. A better measure of the difference comes from comparing directories with distribution areas that perfectly overlap. Bell Atlantic and R.H. Donnelly both have directories with distribution areas that are exactly equal to the boundaries of Washington D.C. The phone company's (Bell Atlantic's) directory is 1,443 pages long, charges $3,387 for a double-quarter column advertisement and collects 7.6 references per household per month. In contrast, the Donnelly directory is 947 pages, charges $2,352 for the same size advertisement and collects 1.4 references per household per month. Despite these numbers, some independent publishers are successful. About 25% of the directories published by telephone companies face a competitor that takes 25% of the usage market or more. In summary, telephone directories company directories are very successful but face non-trivial competition from independents.

Two assumptions about the behavior of publishers follow from *a priori* observation of the industry. First, although Yellow Pages publishing is dominated by regulated utilities, this paper models publishers as profit maximizers. Publishers vigorously protect Yellow Pages profits, often transferring them to for-profit subsidiaries or otherwise hiding profits (White and Sheehan, 1992). The second assumption is that the number of books distributed is exogenous to the price- and quantity setting process. Telephone companies and independents distribute one directory to every phone line in an area, so the number of directories is determined by the geographic scope of a directory. Because scoping is determined before the sale of advertising, I take population coverage as exogenous. [4]

3.2 Data

The data comes from a number of independent sources. National Yellow Pages Monitor (NYPM), a proprietary data company, collects usage data for individual directories. At the request of a client (normally a large Yellow Pages publisher), NYPM measures the number of references per household per month going to all directories in a given metropolitan statistical area, taking each directory area as a unit of observation. [5] The NYPM set contains data on 476 directories.

The number of pages in a directory proxies for the quantity of advertising. The Yellow Pages Publishers Association (YPPA), an industry trade group, maintains a library of directories published by members and the Boston Consulting Group collected the number of pages in each directory for a separate project. The data on page numbers includes a check for directories that are observably small. I convert pages into a variable called "advertising" by multiplying

[4] Telephone companies are required to distribute one White Pages directory to every phone line and most choose to publish the Yellow Pages in the same book as the White Pages. Independent publishers match telephone companies in both ways.

[5] NYPM survey respondents maintain diaries of their Yellow Pages usage for 1 week. NYPM normally surveys 1,000 to 3,000 people per MSA, although NYPM uses 11,200 respondents in Los Angeles, normally leading to a few hundred respondents for even relatively small directories.

pages by the number of columns and multiplying by an adjustment (0.8) for directories that have small height and widths. YPPA membership covers over 6,500 directories including almost every directory in the United States. The YPPA claims that its membership accounts for 95% of Yellow Pages advertising sales.

Advertising prices come from the *Rate and Data* data set and directory statistics come from the *Industrial Characteristics* data set, both from the YPPA. The pricing data is an especially rich source. The *Rate and Data* set contains prices for every size and color of advertisement at every directory in the YPPA. Directories in the data set offer an average of around 80 choices of advertisement size and color. Unfortunately, there is only one observation on quantity to match with each set of prices. In estimation, I choose one price to represent each directory. The *Industrial Characteristics* data set contains directory information such as the number of people a directory covers, and the number of columns in a directory.

Directory boundary data come from Claritas, Inc, a proprietary data company. The boundary data come in the form of computer maps that I matched with population centers of 5-digit zip codes. I assume that if a population center of a zip code falls within the boundary of a directory's distribution area, then the publisher distributes the directory to the entire zip code. Zip codes are a reasonably close approximation of directory boundaries, and in some cases coincide exactly. Combining population data at the zip code level with the mapping data determines the choice set of directories for consumers. All data are from 1996 except for the boundary data which are from 1997. This difference creates some discrepancies and, after matching all four data sets together, the data set contains 428 observations. [6]

I assign directories to their central or most populous counties and match them to demographic data from the *USA Counties CD ROM 1996*. Industry sources suggest that educated, relatively wealthy people who own their own home are likely to use the Yellow Pages, as are people who have recently moved. People who live in urban settings or regularly use public transportation use Yellow Pages less. Table 4 in the Appendix presents a description of

[6] I removed directories that were targeted for non-English speaking audiences, such as Spanish or Chinese language directories.

demographic data that I use to capture these features.

Table 1 presents simple statistics for all of the variables. Note that there is a strong positive correlation between usage and the quantity of advertising, suggesting a network effect. [7] Also, there is a strong negative correlation between measures of competition and price, advertising and usage. Table 1 provides two measures of competition: "ads at competing books" is the total amount of advertising in competing books, averaged over households in a directory's distribution area; and "usage at competing books" is the average number of references per household in a directory's distribution area that go to a competing book. Competing books are those published by other publishers. These negative correlations are robust to controlling for whether or not a directory is associated with a telephone company, and to controlling for demographics.

4 Model

This section presents a model of competition in the Yellow Pages industry suitable for estimation. The model explicitly captures the interaction between directory advertising and directory usage. The model is a one period, simultaneous move, quantity-setting game. [8] There are J publishers that each produce one directory. The distribution areas are taken to be exogenous. The distribution areas may or may not overlap, and may do so for only a portion of their areas. The publishers face two interacting markets: a market for advertising and a market for consumer usage. Consumers receive directories for free so there are no profits directly from the usage market. But the amount of consumer usage affects demand in the advertising

[7] There is already some evidence that consumers value Yellow Pages for the advertising. Laband (1986) and Mixon (1995) show that Yellow Pages advertisements are likely to be more informative if they are for products that are "search goods" (i.e. goods that are expensive and are purchased infrequently). Equally as convincing is a recent Ameritech radio advertisement boasting that the Ameritech Yellow Pages have "the most ads and the most complete information."

[8] Estimating a price-setting game instead of a quantity-setting game presents a number of difficulties, although I still find empirically that a positive feedback loop exists in that case. Section 4.3 discusses these issues.

Table 1: Simple Statistics

	Usage	Price (DQC)	Pages	Advertising*	Population Coverage	Telco Dummy
MEAN	4.85	1787	630	2621	386	0.64
St. Dev.	4.17	1155	479	2190	411	0.48
Corellations						
Usage	1.00	0.37	0.46	0.52	0.08	0.62
Rate (DQC)		1.00	0.73	0.68	0.76	0.33
Pages			1.00	0.98	0.71	0.33
Advertising				1.00	0.66	0.41
Pop Cov					1.00	0.06
Telco						1.00
Ads at competing books	-0.64	-0.28	-0.27	-0.31	-0.06	-0.35
Usage of competing books	-0.61	-0.38	-0.36	-0.41	-0.12	-0.82
% urban population	-0.14	0.14	0.16	0.17	0.24	-0.04
% have not moved	0.16	0.22	-0.03	-0.04	0.04	0.11
% lived in diff county	0.19	-0.11	-0.05	-0.02	-0.20	0.09
% lived in diff state	-0.09	0.08	0.23	0.21	0.17	-0.07
% take public trans.	-0.01	0.30	0.21	0.20	0.27	0.00
% own house	0.20	0.01	-0.06	-0.08	-0.18	0.15
% grad hi school	-0.04	0.08	0.12	0.10	0.01	0.04
% grad college	-0.11	0.17	0.17	0.15	0.11	0.01
establishments per cap.	0.04	0.25	0.18	0.16	0.12	0.07
per cap income	0.02	0.22	0.16	0.16	0.08	0.06
house-building permits	-0.01	0.05	0.02	-0.02	-0.07	0.10
county pop. growth rate	-0.04	-0.13	-0.08	-0.10	-0.16	0.01
pop. density	-0.08	0.23	0.15	0.15	0.27	-0.03
earnings per worker	-0.09	0.27	0.19	0.17	0.21	0.00
GTE	-0.05	-0.09	-0.06	-0.01	-0.09	0.02
Bell South	0.22	0.17	0.07	0.05	-0.01	0.18
Distribution Area	0.09	0.01	0.06	0.06	0.08	0.01
Rate (Full Page)	0.57	0.91	0.80	0.81	0.71	0.52
Rate (Bold)	0.44	0.90	0.66	0.63	0.64	0.38

* Advertising equal Pages multiplies by columns and a size adjustment for small directories.

market. [9] Specifically, each publisher j, $j = 1, ..., J$, faces two demand curves: retailer inverse demand for advertising $P_j(A_1, U_1, ..., A_J, U_J)$ and consumer demand for usage $U_j(A_1, ..., A_J)$, where A_j is the amount of advertising at j and U_j is the number of uses per consumer covered.

I expect that $\partial P_j/\partial U_j > 0$ and $\partial U_j/\partial A_j > 0$. These two conditions together represent the network effect. If usage increases for some exogenous reason, then the inverse demand curve for advertising shifts out because $\partial P_j/\partial U_j > 0$. In equilibrium, the shift leads to an increase in the quantity of advertising that then implies a further increase in usage because $\partial U_j/\partial A_j > 0$. The positive feedback loop between usage and advertising means that an otherwise small change in either demand curve can lead to a large change in both usage and advertising.

I further expect that $\partial P_j/\partial A_j < 0$, representing the "scarcity" effect, the standard effect that price decreases in quantity. I also expect that $\partial U_j/\partial A_k \leq 0, k \neq j$. I discuss $\partial P_j/\partial A_k$ and $\partial P_j/\partial U_k$ in Section 4.1. Each publisher simultaneously chooses its quantity of advertising A_j to sell in its directory. [10] The solution concept is Nash equilibrium.

In summary, there are two effects from an increase in advertising. An increase in advertising leads to a price decrease because it is a movement along the willingness-to-pay curve. However, usage increases which shifts the willingness-to-pay curve out. The locus of points that account for both changes is the demand curve that the publisher faces. Willingness-to-pay curves are demand curves with usage held constant – the slope of a willingness-to-pay curve is measured by $\frac{\partial P_j}{\partial A_j}$. The slope of the demand curve that the publisher faces is $dP_j/dA_j = \frac{\partial P_j}{\partial A_j} + \frac{\partial P_j}{\partial U_j}\frac{\partial U_j}{\partial A_j}$. [11]

Although most of the theoretical results that follow hold for general demand functions, Sections 4.1 and 4.2 present functional forms for usage and advertising designed for purposes of estimation. Sections 4.3 and 4.4 discuss equilibrium and efficiency, and 4.4 presents a measure

[9] Presumably, if publishers did charge for directories, the same issues would apply. That is, a low price for directories would increase the demand for advertising, which would then increase consumer demand for directories.

[10] In the data, some publishers own overlapping directories. I address this feature in the construction of the publishers' first-order conditions.

[11] Economides and Himmelberg (1995) make a similar distinction and point out that in many cases, network effects imply an upward sloping demand curve for low levels of quantity. Under strong network effects, no one will pay for a product that no one else uses.

of forgone surplus due to a failure to take advantage of network effects.

4.1 Demand for Advertising

The functional form for $P_j(.)$ is driven by surprising results from estimation. The function $P_j(.)$ clearly should depend on advertising and usage at book j, as well as demographic variables. One might think that it should also depend on advertising and usage at competing directories. But I estimate these coefficients to be close to zero and insignificant. Section 4.1.1 presents a model of individual advertiser demand that generates this result and aggregates to the functional form for $P(.)$, thereby allowing for a structural interpretation of the parameters in $P(.)$. Section 4.1.2 discusses empirical identification.

4.1.1 A Model of the Advertising Market

A representative advertiser chooses the amount of advertising a_j (a continuous choice variable) to place in directory j. The advertiser acts as a price taker and chooses its optimal level of advertising, $a_j(P_1, ..., P_J)$. The market size of advertising is $\bar{m}$, so total advertising in book j is $\bar{m}a_j = A_j$. By inverting the aggregate demand curve, we can construct an inverse demand curve that the publisher faces.

In order to derive the factor demand for advertising $a_j(P_1, ..., P_J)$, I use the following set-up: Each consumer needs information some exogenous number M times per month. Each time they need information, they can use a Yellow Pages directory in their area or some other option, such as the internet or word-of-mouth. [12] I expect that a consumer is more likely to use a Yellow Pages directory when it is more informative, or equivalently, when it has more advertising. The first assumption on consumer behavior is:

[12] The parameter M captures something fairly abstract. For instance, every time a person needs a haircut, they have an information requirement in the sense that they need to pick a hair salon. In order to choose, they could look in the Yellow Pages or they could use an "outside option" such as going to the place they went to the last time, getting a recommendation from a friend or going into a salon they have seen in their neighborhood.

A1 Consumers use at most one directory per information requirement.

Rochet and Tirole (2002) refer to this feature as "single-homing." This assumption would naturally follow from a scenario in which there was some cost to opening a second directory or storing it in an accessible place, especially if consumers found directories to be close substitutes. NYPM does not keep track of the simultaneous use of multiple directories explicitly, but this conforms with casual observation as well as some related statistics. For instance, when consumers reference a directory, they contact an advertiser 82.1% of the time, but make more than one contact only 36.2% of the time. These data fit with consumers who use only one directory, although are not definitive. Note that under this assumption, consumers may use different books for different needs. For instance, they may use a local book when they need a barber but a regional book when they need a car dealer. Assumption $A1$ says that consumers do not open up multiple directories simultaneously. Anderson and Coate (2001) make a similar assumption. As in their work, this assumption implies that directories are monopolists over access to their readers.

Once a consumer has opened a directory, the consumer looks at a number of advertisements and makes contact with a portion of the advertisers. Advertisers are interested in how many looks are generated from an advertisement. Let the number of times the average person in directory j's distribution area looks at the representative advertisement in book j be $L_j = L(a_j, U_j, A_j)$, where $\partial L_j/\partial a_j > 0$, $\partial L_j/\partial U_j > 0$ and $\partial L_j/\partial A_j < 0$. The last inequality captures the fact that a given advertisement is less likely to be seen in a large book. Note that the L function is identical for each directory, although directories differ in usage and advertising. A proportion of the consumers who look at an advertisement make contact with the advertiser, and these consumer contacts generate some level of profit. The second assumption is:

A2 Advertiser profit per look is constant.

These two assumptions imply that profit is separable in a_j. Intuitively, the first assumption implies that advertising in one book is not a substitute or a complement for advertising in another book. As a result, there is no demand-side reason why the choice of advertising at one book should affect the choice at another book. The second assumption says that having many customers as a result of one advertisement does not affect the cost or benefit of serving customers generated by another advertisement. So there is no cost-side reason why the choice of advertising at one book affects the choice at another book. Separability of the profit function follows. The advertiser's profit function Π can be written as:

$$\Pi = \hat{\pi}_1 L(a_1, U_1, A_1) - P_1 a_1 + ... + \hat{\pi}_J L(a_J, U_J, A_J) - P_J a_J.$$

The term $\hat{\pi}_j$ captures the profit to the advertiser from the number of looks per person received from book j's distribution area. The term $\hat{\pi}_j$ captures variables such as the number of people in the distribution area and their demographics.

Separability implies the desired result that outcomes at each book do not affect each other directly. The rest of this sub-section develops functional forms for estimation. Let $L(a_j, U_j, A_j)$ have the Cobb-Douglas form, so $L_j = a_j^{\gamma_1} A_j^{\gamma_2} U_j^{\alpha_1}$. The parameter γ_1 is expected to lie between 0 and 1, and captures decreasing returns to large advertisements. I expect that the parameter γ_2 will be negative capturing the business stealing effect, i.e. the fact that an advertisement might get lost in a large directory. The parameter α_1 should be positive because more usage of a directory increases the likelihood of consumers looking at a given advertisement. [13]

Now the profit function can be written as:

$$\Pi = \hat{\pi}_1 a_1^{\gamma_1} A_1^{\gamma_2} U_1^{\alpha_1} - P_1 a_1 + ... + \hat{\pi}_J a_J^{\gamma_1} A_J^{\gamma_2} U_J^{\alpha_1} - P_J a_J.$$

The advertiser picks a_j to maximize $\hat{\pi}_j a_j^{\gamma_1} A_j^{\gamma_2} U_j^{\alpha_1} - P_j a_j$. The representative advertiser is too small to affect A_j and takes it as given. The optimal a_j for the advertiser is:

$$a_j = \left(\frac{P_j}{\gamma_1 \hat{\pi}_j A_j^{\gamma_2} U_j^{\alpha_1}} \right)^{\frac{1}{\gamma_1 - 1}}.$$

[13] Note that it is straightforward to obtain these results with advertiser heterogeneity. Let there be a continuum of advertisers indexed by $l \in [0, \bar{m}]$ distributed $f(l)$. Denote the choice of advertiser l at book j as a_{jl}, so $A_j(P_1, ..., P_J) = \int_0^{\bar{m}} a_{jl}(P_1, ..., P_J) f(l) dl$. Instead of $\hat{\pi}_j$, let each $L()$ function be augmented by $\hat{\pi}_{jl}$ and let $\pi_j = (1/\int_0^{\bar{m}} (1/\hat{\pi}_{jl})^{1/(\gamma_1 - 1)} f(l) dl)^{\gamma_1 - 1}$. This change of structure leads to Equation 1. The interpretation of $\hat{\pi}_{jl}$ is of factors idiosyncratic between the directory and the advertiser, such as the location of the advertiser within the directory's distribution area, and π_j is an aggregate of these effects.

Aggregating, we have:

$$A_j = \left(\frac{P_j}{\gamma_1 \pi_j A_j^{\gamma_2} U_j^{\alpha_1}}\right)^{\frac{1}{\gamma_1 - 1}}$$

where $\pi_j = \hat{\pi}_j / \bar{m}^{\gamma_1 - 1}$ is an aggregate equivalent of $\hat{\pi}_j$ that accounts for the mass of advertisers in the market. Solving for P_j, we obtain the inverse demand curve:

$$P_j(A_j, U_j) = \gamma_1 A_j^{\gamma_1 + \gamma_2 - 1} U_j^{\alpha_1} \pi_j. \tag{1}$$

There are several important features of this demand curve. First, it should increase in usage (I expect $\alpha_1 > 0$). This feature, along with the fact that usage increases in advertising, represents the network effect. Second, the demand curve decreases in advertising both because there are decreasing returns to individual advertisers from large ads (I expect $\gamma_1 < 1$) and because advertisements will be lost in a large book (I expect $\gamma_2 < 0$), but these effects will not be distinguishable in estimation. In estimation, I refer only to estimating γ, where $\gamma = \gamma_1 + \gamma_2 - 1$.

Third, the price and quantity of advertising and the amount of usage in another book do not affect book j directly. Holding usage at book j constant, advertisers are willing to pay the same amount to advertise in book j regardless of what book k does. This result follows from Assumptions *A1* and *A2*. If consumers chose books at random so the consumer side of the network effect is eliminated, then publishers would have significant market power as there is no other way to reach these consumers. But with the network effect, book k competes with book j to attract usage. The inverse demand curve $P_j(A_1, ..., A_J, U_1(A_1, ..., A_J), ..., U_J(A_1, ..., A_J))$ can be rewritten as $P_j(A_j, U_j(A_1, ..., A_J))$. This feature is consistent with estimation results and, as it is difficult to construct a model of how advertisers trade off between directories, makes a structural interpretation of parameters much easier.

4.1.2 Identification

A formal discussion of the estimation is delayed until Section 5. Here we discuss intuitively the identification of the parameters in Equation 1. In practice, I specify:

$$\ln(P_j) = \gamma \ln(A_j) + \alpha_1 \ln(U_j) + X_j^P \beta^P + \nu_j. \tag{2}$$

That is, $\ln(\pi_j)$ is captured by a linear function of observable variables and an unobservable term ν_j. As the equilibrium quantity of advertising depends on price, we expect $\ln(A_j)$ and $\ln(U_j)$ to be correlated with ν_j. For instance, if willingness-to-pay was high for unobservable reasons, we might also expect the quantity of advertising to be high via the publisher's first-order condition, and therefore usage to be high as well.

I address this problem with instrumental variables. As an instrument for usage, I use variables that capture the number of people who recently moved. People who recently moved tend to use Yellow Pages much more then long-time residents. This instrument works well as long as recent movers do not tend to be more valuable customers – that is, recent movers do not affect the demand for advertising over and above their effect on usage. [14] In practice to identify α_1, I use the percentage of people in the county who lived in a different state 5 years previous, the percentage who lived in different county 5 years previous and the percentage who lived in the same house 5 years previous.

In order to identify the effect of the quantity of advertising, I use variables that would be expected to move marginal cost and therefore affect the publisher's first-order condition. Almost all publishers contract publishing to a single firm, R.R. Donnelly. However, Bell South and GTE maintained their own printing facilities. I use dummy variables for being one of these companies as instruments for γ. Wages at the level of the publisher would make excellent instruments but are unobservable. Instead I use the census measure of earnings level in a county,

[14] For instance, if recent movers are more likely to be forming long-term relationships with retailers, then recent movers might be valuable over and above their effect on usage. But in practice, most users of Yellow Pages are forming long-term relationships. Consumers who already have retailers typically have the number or use the White Pages.

which approximates a county-level average hourly wage. Local wages are useful because the biggest cost to producing a directory is sales, and salespeople must be based near to the county where the directory distributes. Local wages are problematic to the extent that they also affect advertising demand - I control for this issue by including county level income in X_j^P. Section 5 discusses the effectiveness of these instruments.

4.2 Consumer Usage

In the model of the advertiser demand, I assume that consumers choose one directory or an outside option each time they need information. That naturally suggests a discrete choice model for the consumer choice of a directory. Once a consumer picks a directory, they may look at any number of advertisements – their behavior generates the look function $L()$ discussed above. But instead of formally modeling how consumers choose among advertisers, I simply allow the utility from choosing a directory to depend on total advertising in a log-linear fashion. This simplification is likely to be reasonably accurate and is difficult to improve upon without more disaggregate data. Section 4.2.1 presents the discrete choice model and Section 4.2.2 discusses identification.

4.2.1 Model

I follow methods presented in Berry (1994) for applying a nested logit model to markets with aggregate data. As above, the total number of times that a representative household requires information of the kind that can be found in the Yellow Pages is an exogenous number M. In order to allow Yellow Pages directories to be closer substitutes with each other then with the outside option, I place Yellow Pages directories in one nest and the outside option in a separate nest. Following Berry (1994), let the utility to consumer i from directory j be:

$$u_{ij} = \alpha_2 \ln(A_j) + X_j^U \beta^U + \xi_j + \zeta_i(\sigma) + (1-\sigma)\epsilon_{ij}.$$

I expect to find that α_2 is positive, which along with α_1 captures the network effect. The vector X_j^U contains demographic variables associated with the central county of each directory. The variable ξ_j is a directory-specific variable that captures characteristics that are unobservable to the econometrician. In this case, the characteristics could be unobservable quality of the directory or region-specific usage effects. The unobservable variable $\zeta_i(\sigma)$ captures individual i's preference for Yellow Pages and ϵ_{ij} captures individual preference for a specific directory. The variable ϵ_{ij} captures issues such as the location of the consumer relative to the location of the directory. I assume that ϵ_{ij} is distributed Type I Extreme Value and $\zeta_i(\sigma)$ is distributed such that $(1-\sigma)\epsilon_{ij} + \zeta_i(\sigma)$ is also distributed Type I Extreme Value. Berry (1994) discusses this issue and Cardell (1997) shows that ζ_i exists and is unique. The parameter σ is restricted to lie between 0 and 1, and measures the correlation in unobserved (to the researcher) utility from different Yellow Pages directories. As σ approaches 0, correlation within the group goes to zero and the model approaches a standard logit model. The parameter σ will be estimated. The utility from the outside good is normalized to be $u_{ij} = \zeta_{i0}(\sigma) + (1-\sigma)\epsilon_{i0}$.

Let s_j be the market share to directory j in its distribution region, so $U_j = Ms_j$. Let $s_{j|YP}$ be the share of people who choose directory j given that they choose to use a Yellow Pages directory and let s_0 be the share to the outside option in j's region (the j is suppressed as it will be obvious in context). It is useful to define the mean utility for directory j to be δ_j, so

$$\delta_j = \alpha_2 \ln(A_j) + X_j^U \beta^U + \xi_j. \tag{3}$$

Berry (1994) shows that under the nested logit assumptions, we have

$$s_j = e^{\delta_j} s_0 s_{j|YP}^{\sigma}.$$

This relation allows for the identification of α_2, β^U and σ in a log-linear model. If this relation held true at every observation, I could estimate α_2, β^U and σ by:

$$\ln(s_j) - \ln(s_0) = \alpha_2 \ln(A_j) + X_j^U \beta^U + \sigma \ln(s_{j|YP}) + \xi_j. \tag{4}$$

Note that we observe $s_{j|YP}$ in the data and compute s_j and s_0 by making an assumption about M, the total number of information requirements. Assuming that people need information the same amount of times seems an appropriate first approximation. An alternative would be to allow M to depend on demographics, although that would be asking a lot of aggregate data. Instead, I try different values of M and show that the value does not significantly affect results.

A comment is in order on the assumption that ϵ_{ij} is *iid*. We expect the preferences of households to be correlated across time and the average number of choices per households per month ranges as high as 25, so the *iid* assumption could be problematic. However, less restrictive assumptions would suffice. For instance, allowing for correlation in ϵ_{ij} within households but assuming ϵ_{ij} is uncorrelated with the number of times a household needs information is sufficient. To see this, take the extreme case where each household chooses M times and draws the same ϵ_{ij} each time. Households choose the same directory each time but regardless of whether M is 1, 5 or 30, market shares remain the same and so Equation 4 holds. As long as households that prefer a particular book do not also need information more or less often, we may proceed using Equation 4.

There is a separate reason why Equation 4 is problematic. Equation 4 can only be applied to areas where all consumers have the same choice set whereas the market for Yellow Pages is characterized by overlapping markets with distinct boundaries. [15] In principle, one could apply (4) to each area with a uniform set of directories, but the data set contains usage shares for directories only in their whole area. One cannot tell how much usage of a directory comes from the area where the directory is a monopolist as opposed to from the area where the directory faces competition. Fortunately, it is possible to infer what usage must be in each sub-market by combining data on total usage share, data on the extent of overlap and the assumptions of the nested logit model. [16]

Briefly, usage shares for each sub-market can be constructed based on the nested logit model

[15]The term "overlapping markets with distinct boundaries" distinguishes the problem from more standard problems of market areas. For instance, two restaurants that are distant from each other may have common customers but they do not have distinct boundaries. All consumers could choose to go to either restaurant. The problem in this paper is one in which different groups of consumers have different choice sets. Consumers cannot use a directory unless they live in the directory's distribution area.

[16]Sub-markets are defined as areas served by a uniform set of directories. In this data, 428 directories generate 660 sub-markets, with some sub-markets being served by as many as 8 directories.

from a vector δ and the parameter σ. I develop an algorithm that finds the vector δ that implies sub-market usage shares that add up to observed market shares. While there is no explicit function for δ, it is straightforward to nest a fixed-point algorithm into an optimization routine as suggested by Rust (1987) and Berry, Levinsohn and Pakes (1995). With δ in hand, we can estimate the desired parameters from Equation 3. In order to increase identification power over σ, I use Equation 4 at the 56 observations where it is appropriate, i.e., the directories that contain only 1 sub-market because they are completely overlapped by all competing directories. Details of this routine appear in the Appendix.

4.2.2 Identification

Equation 4 introduces two identification problems. [17] The first is the same as discussed before, usage share and advertising are determined simultaneously. As an instrument, I use the number of people covered by a directory. Population coverage should not affect a household's decision about which directory to choose but should have a major impact on the demand for advertising. That is, population coverage appears in X_j^P but not X_j^U and so is an appropriate instrument.

Another issue is to identify σ. Conceptually, the share of consumers switching to the Yellow Pages group as the features of the group change identifies σ. Such variation can be a result of either changes in directory characteristics or changes in the number of directories. However, there is a potential endogeneity problem with this second type of identification if directories are attracted to markets where usage is high for reasons that are unobservable to the researcher. In Equation 4, if ξ_j is high, s_j will be high and there will be more entrants so $s_{j|YP}$ will be low. The estimate of σ will be biased downwards unless $s_{j|YP}$ is properly instrumented for. For this purpose, I use the square mileage of the distribution area of a directory as an instrument for the usage equation. An empirical fact is that larger directories have less of their region overlapped by other directories and so should have a higher within-group share. The simple correlation between the number of directories in an area (weighted by population) and the size of the

[17] The result of the fixed point algorithm described at the end of Section 4.2 has the same identification problems.

show the positive feedback loop. For this reason, I only apply enough instruments to just identify the first-order condition so it does not impact the estimates of the demand parameters, only their standard errors. [18] Indeed, when I estimate the demand functions without the first-order condition, I get almost identical parameter values.

Before moving on, I briefly discuss the use of a quantity-setting model instead of a price-setting model. Estimating the first-order condition in the price-setting game introduces serious difficulties because usage depends directly on quantity. The primary problem is a computational one. A price-setting game would require specifying a demand curve of the form $A = A(P, U(A))$. The quantity of advertising shows up on both sides of the equation. Therefore, taking the derivative of the demand curve with respect to price (in order to compute marginal revenue) would require solving a fixed-point equation, further complicating an already involved optimization routine. [19]

I circumvent this problem by estimating a quantity-setting game. Note that the difficulty with the price-setting game is only in the estimation of the first-order condition. As I am primarily interested in the demand curves, I also tried estimating the advertiser demand equation with price on the right-hand side and quantity on the left-hand side, which would be a first step towards estimating a price-setting game. When I do so, I get very similar results to those reported here.

[18] In this sense, the first-order condition is just-identified. This approach differs from Berry, Levinsohn and Pakes (1995) where part of the identification of demand parameters comes from over-identifying the first-order condition.

[19] A secondary problem involves multiple equilibria. Conditional on directory choices, a price-setting game admits multiple equilibria whereas a quantity-setting game does not. Consider two symmetric directories. If they both choose the same price, it is reasonable to expect multiple equilibria where one directory has high advertising and high usage while the other has low advertising and low usage. In this case, determining marginal revenue from a change in price requires an arbitrary assumption about which equilibrium is selected. Conversely, if two symmetric directories pick the same quantity, then they have the same usage and therefore the same price. The quantity-setting game can still have multiple equilibria but not conditional on quantities. Therefore, the first-order condition in the quantity-setting game, which takes competitors' choices as given, is straightforward. In fact, for the parameters estimated in this paper, there is only one equilibrium.

4.4 Efficiency

Consumers obtain directories for free and consumers do not have the option to pay for a directory with extra advertisements. The lack of prices for the consumer side of the market makes it difficult to convert consumer utility into units that are easily comparable to retailer surplus. I analyze surplus formally only from the point of view of advertisers but I discuss consumer surplus when I am able to do so. [20]

A social planner would choose the set of advertising levels A_j, $j = 1, ..., J$, simultaneously to maximize $\sum_{j=1}^{J} \int_0^{A_j} P_j(s, U_j(A_1, ..., A_J))ds - C(A_j)$. The first-order conditions for the social planner are:

$$P_j + \sum_{k=1}^{J} \int_0^{A_k} \frac{\partial P_k(s, U_k(A_1, ..., A_J))}{\partial U_k} \frac{\partial U_k}{\partial A_j} ds - MC_j = 0 \quad j = 1, ..., J. \tag{6}$$

It will be rare for Equations 5 and 6 to hold simultaneously. [21] It cannot be said for sure which regime will result in more advertising. The social planner accounts for the value of the network effect to the entire set of advertisers and the advertisers in other directories. The publisher takes into account the value of the network effect on its marginal purchaser. However, the publisher also takes into account the effect of downward sloping demand on marginal revenue and in most reasonable cases, we have the standard result that the size of the network is too small. By "reasonable", I mean that the network effect of advertising on a publisher's own price is stronger than the negative effect on its competitors' prices, so $\sum_{k=1}^{J} \frac{\partial P_k}{\partial U_k} \frac{\partial U_k}{\partial A_j} > 0 \; \forall A_j, \; j = 1, ..., J$. In this case, the social planner picks A_j such that price is less than marginal cost. "Reasonable" also means that demand is downward sloping, even accounting for the network effect, so $\frac{\partial P_j}{\partial A_j} + \frac{\partial P_j}{\partial U_j} \frac{\partial U_j}{\partial A_j} < 0$ at the Nash equilibrium, which implies that price is greater than marginal cost in the Nash equilibrium. These conditions are testable and I return to them in Section 6.

[20] Berry and Waldfogel (1999) take a similar approach.

[21] I have assumed that the publishers can set only one price. If a publisher can first-degree price discriminate, it could achieve the socially optimal level of advertising. This result is the case in Liebowitz and Margolis (1994) who assume that a monopolist faces identical purchasers, which implies that the monopolist can charge consumers their valuation with a single price.

In equilibrium, some of the forgone surplus is a result of publishers exercising market power while some is due to the market's failure to account for the effects of aggregate advertising on the profits of individual advertisers. This paper uses the following definitions to differentiate between these two effects. Let *classical deadweight loss* be the difference between the amount of surplus generated at the Nash equilibrium and the amount a social planner would generate if it did not account for the network effect. Let *network deadweight loss* be the difference between the amount of surplus generated when the social planner accounts for the network effect and when it does not. The following discussion defines these terms formally and Figure 1 demonstrates. Fix the competitors' outcomes at their Nash equilibrium choices A_{-j} and denote the equilibrium choice of advertising for publisher j as A_e. Without network effects, the social planner would take the demand curve to be the willingness-to-pay curve $P_j(A_j, U_j(A_e, A_{-j}))$. Denote the social planner's choice of advertising level when usage is fixed at $U(A_e, A_{-j})$ as A_0. Let $(A_0 - A_e)MC_j$ be the cost of the extra production, assuming constant marginal cost (as is assumed in estimation). Classical deadweight loss is:

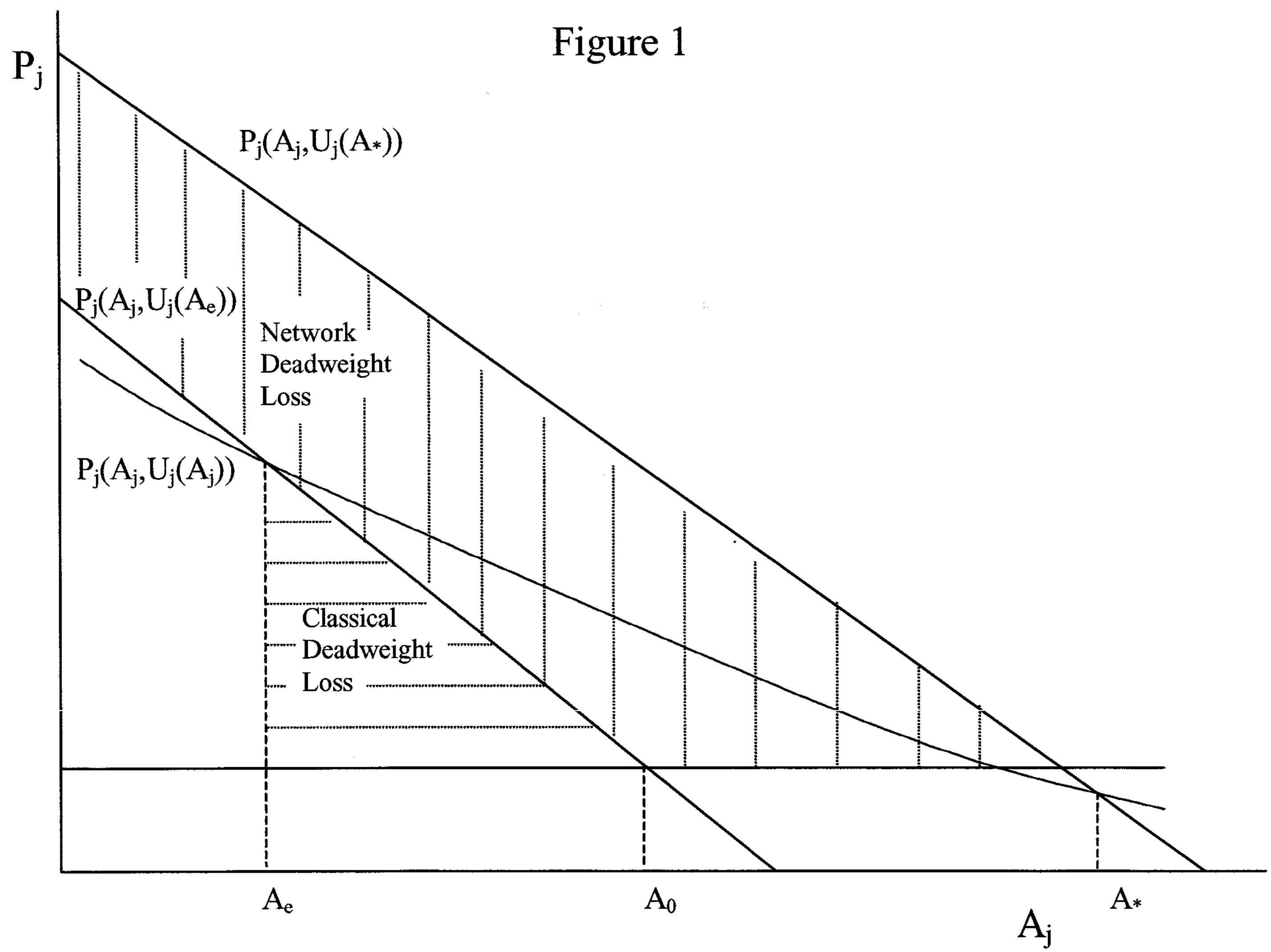
$$\int_{A_e}^{A_0} P_j(s, U_j(A_e, A_{-j}))ds - (A_0 - A_e)MC_j.$$

If a social planner raised advertising from A_e to A_0, usage would rise above $U(A_e, A_{-j})$, which means that advertising would generate more surplus, so the efficient level of advertising would be even higher. Denote the socially efficient level of advertising by firm j and its competitors as A_* and A_{-j*}. Network deadweight loss is defined as:

$$\int_0^{A_*} P_j(s, U_j(A_*, A_{-j*}))ds - \int_0^{A_0} P_j(s, U_j(A_e, A_{-j}))ds. - C_j(A_* - A_0).$$

In the figure, the space between the efficient willingness-to-pay curve and the equilibrium willingness-to-pay curve is network deadweight loss. As stated previously, it will often be efficient for the willingness-to-pay of the last purchaser to be below marginal cost because of the network effect. Estimates of structural parameters will allow us to measure network deadweight loss, classical deadweight loss and equilibrium consumer surplus. Clearly, it is important to use a

Figure 1



P_j
$P_j(A_j,U_j(A_*))$
$P_j(A_j,U_j(A_e))$
Network
Deadweight
Loss
$P_j(A_j,U_j(A_j))$
Classical
Deadweight
Loss
A_e
A_0
A_*
A_j

model which distinguishes between $\frac{\partial P}{\partial A}$ and $\frac{\partial P}{\partial A} + \frac{\partial P}{\partial U}\frac{\partial U}{\partial A}$ in order to perform this calculation. [22]

5 Estimation

To review, we estimate 3 equations simultaneously, Equations 2, 4 and 5. [23] While the derivation of the functional forms for estimation in Section 4 is involved, the final functional forms are quite intuitive. Equation 2 is price regressed on a log-linear function of advertising usage and demographics and Equation 4 is usage share regressed on advertising and demographics, with some adjustments so it can be interpreted as a nested logit model.

5.1 Methodology

Because the presence of demand parameters in marginal revenue creates non-linear cross equation restrictions on parameters, I estimate this system simultaneously using the Generalized Method of Moments (Hansen 1982). For each equation, I construct a matrix of exogenous variables Z^P, Z^U and Z^C, and assume $\mathsf{E}([\nu\ \xi\ \omega]|Z^P, Z^U, Z^C) = 0$. Each Z matrix contains exogenous variables in its respective equation (i.e. X^P, X^U, and X^C), as well as the "excluded" variables discussed in Sections 4.1.2 and 4.2.2. Descriptions of the variables can be found in Table 4 in the Appendix. The estimation problem is to choose parameters $[\alpha\ \beta\ \gamma\ \sigma]$ to minimize the criterion $[m'\Phi m]$, where Φ is a positive definite weighting matrix and:

$$m = \begin{bmatrix} Z^{P'}\hat{\nu} \\ Z^{U'}\hat{\xi} \\ Z^{C'}\hat{\omega} \end{bmatrix}$$

[22]Another reasonable choice for the definition of A_o would be the level of advertising A_j that sets demand $P_j(A_j, U(A_j, A_{-j}))$ equal to marginal cost. This choice has some appeal but using $P_j(A_j, U(A_j, A_{-j}))$ clearly involves network effects. In order to isolate network effects as much as possible, I define A_o with the willingness-to-pay curve $P_j(A_j, U(A_e, A_{-j}))$, so usage is held constant.

[23]Equation 4 is for the case where a directory is overlapped by all competitors. In other cases, Equation 4 is replaced with the calculation discussed at the end of Section 4.2.

The $(J \times 1)$ matrices $\hat{\nu}$, $\hat{\xi}$ and $\hat{\omega}$ are estimates of ν, ξ and ω based on estimates of α, β, γ and σ. Hansen (1982) shows that this estimator is consistent for any positive definite Φ, and that the estimator is efficient if Φ is chosen to be the inverse of the correlation matrix of the vector m.

For price, I use the rate for a double-quarter column advertisement. This rate is the most closely watched rate in the Yellow Pages industry and is available at practically all directories in my sample. When I used other rates, results did not change (note the high correlation between rates in Table 1). For the quantity of advertisements, I used the number of pages multiplied by the number of columns. I multiplied this number by 0.8 if the book was observably smaller than a standard directory. I focus on the results for $M = 75$, although I also present results for $M = 35$ to show that the differences are not substantial.

5.2 Instruments

Computationally, parameters on endogenous variables (α_1, α_2, γ, σ) are identified by variables that appear in the corresponding Z^P or Z^U matrix but do not appear in the equation which defines the parameter. These exclusion restrictions are reviewed in Table 2. Table 2 lists instruments next to each endogenous variable. But note that in GMM, instruments are applied to equations, not variables. So computationally, each instrument identifies both endogenous variables in its equation. Equation 5 does not introduce any further endogeneity problems so I use $Z^C = X^C$.

In GMM, there is no equivalent to the first stage of Two Stage Least Squares, but Table 4 replicates the "first stage" in order to gain intuition about identification. Table 4 presents OLS regressions of endogenous variables on their respective exogenous variables. That is, $\ln(A)$ and $\ln(s_{j|YP})$ are regressed on Z^U, and $\ln(A)$ and $\ln(U)$ are regressed on Z^P. Variables in bold represent exclusion restrictions. For the usage equation, advertising increases in population coverage, and within-group share increases in distribution area, both as expected. For the advertising demand equation, the quantity of advertising decreases in earnings and usage

Dependent Variable	Endogenous Variable	Instruments
s_j/s_0	A_j	Population Coverage
	$s_{j\|YP}$	Distribution Area
P_j	A_j	Earnings
		Pub Dummies(GTE, Bell South)
	U_j	% switched county
		% switched state
		% in same house

Table 2: Orthogonality Restrictions

increases in the number of people who lived in a different state and a different county five years previous. The only unexpected result is that usage increases in the number of people who owned the same house as five years previous. The R^2 statistics are fairly high, always greater then 0.5. Also, for each regression, an F-test rejects joint insignificance of the excluded variables at a 95% confidence level. [24]

6 Results

Table 5 presents the main results. The effect of usage on advertising and the effect of advertising on usage are positive and significant, implying the existence of a network effect. [25] The effect of the quantity of advertising on the price of advertising is negative and significant. The estimate of σ is high (0.803), suggesting that consumers view directories as similar products. Note that consumers have much stronger demand for directories associated with

[24] Note that there are two "first-stage" regressions for $\ln(A_j)$. This follows because I use separate sets of instruments for each equation although formally all instruments could be applied to all 3 equations.

[25] If doubling average usage doubled demand for advertising, we would expect α_1 to be 1. There are number of explanations why α_1 is significantly less then 1. It may be that in areas with high usage, consumers are more willing to search through listings and small advertisements, which mitigates consumers' impact on the demand for advertising. Another possibility is that consumers that use directories more use them for different headings, as opposed to using each heading more often. This behavior could have potentially complicated effects on demand. Both of these issues are difficult to address with this data set.

telephone companies. There does not seem to be a similar effect in advertiser demand. Most of the other coefficients are insignificant, although most of the ones that are significant are of the expected sign (for instance, the percent of people who lived in different states and the percent in different counties 5 years previous are positive in usage, as expected). Two surprising results are that the percentage of college graduates and the number of new home building permits are negative in usage. It may be that the opportunity cost of time for college graduates outweighs the effect that higher educated people are more likely to use Yellow Pages. Also, the number of new homes as a percentage of county population is supposed to capture growth but may actually proxy for sparsely populated areas, which often have low Yellow Pages usage.

Table 6 presents 3 other specifications in order to explore the robustness of the results. The first column presents results without instrumenting. In this specification, I hold σ fixed at 0.8 in order to focus on changes in the other coefficients and also to speed convergence. [26] As expected, the coefficient on quantity in price is closer to zero then in Table 5, -0.004 instead of -0.729. Also, the coefficient on advertising in usage (α_2) is higher (0.230 relative to 0.154) when not instrumenting. However, an unexpected result is that when not instrumenting, the coefficient on usage in price (α_1) is lower (0.116 instead of 0.564), suggesting that endogeneity is biasing that coefficient towards zero. One explanation for this result may be that the publisher first-order condition implies that advertising is a concave function of price. Also, usage is endogenous only because it is a (non-linear) function of advertising. That is, the coefficient on usage in the regression without instrumenting may be lower then expected because it is capturing some of advertising's non-linear relationship with price. This argument implies that it is necessary to instrument for both advertising and usage to see the exogenous effect of either variable. In fact, when I instrument for only advertising or only usage, the coefficients do not change nearly as much as when I instrument for both. In a separate regression, I estimate the usage equation without instrumenting for $s_{j|YP}$, the within-group market share. As expected,

[26] Fixing σ fixes δ which means the fixed point may be computed only once at the beginning of the program.

the estimate of σ is much lower, only 0.43 relative to the estimate 0.803 found in Table 5.

The second column of Table 6 presents results when M, the number of times a household needs information per month, is set down to 35. The results are essentially the same except, as expected, the constant term in the usage equation is much higher. The parameter M cannot be set much lower, as I observe areas that use directories up to 25 times per month. Results also do not change when M is set much higher then 75. The third column tests the accuracy of Assumptions *A1* and *A2* by augmenting the vector of explanatory variables for the advertising demand equation with the two "competition variables" from Table 1. Specifically, the demand equation includes an index of the amount of advertising at competing directories and the amount of usage going to competing directories, both in logs. The coefficients are close to zero and insignificant suggesting that outcomes at competing directories do not affect each other's demand for advertising directly. This feature is consistent with advertisers' decisions being separable in their choices at different books, and with Assumptions *A1* and *A2*. Note that competition variables might be endogenously determined and I did not include them in the instrument vector. These results are robust to including the measure of usage at competitors without the measure of advertising at competitors, and vice versa.

Returning to the results in Table 5, the model fits the data well. When I use the 3 equations to predict all 3 endogenous variables simultaneously, the simple correlation between predicted advertising and observed advertising is 0.621. The correlation for price is 0.618 and the correlation for usage is 0.794. Note that the effect of missing a prediction by a small amount in one equation is magnified by the positive feedback structure. As another check, I compute the profits of directories published by telephone companies to be 33.5% of revenue. The number is very close to quotes made to me by members of the industry of 35-45%.

Note that the theoretical model can explain the asymmetric outcomes between directories either with asymmetric directory characteristics or with multiple asymmetric equilibria. In estimation, I find that there is a unique equilibrium (explored in detail in the next subsection) but that directory asymmetries are important, captured by the large coefficient on the dummy variable for being a telephone company directory in the usage equation. Given that telephone directories always seem to be the ones with high price-quantity-usage and independents always have low, it is not surprising that the estimation procedure explained the data with the dummy variable as opposed to finding multiple equilibria.

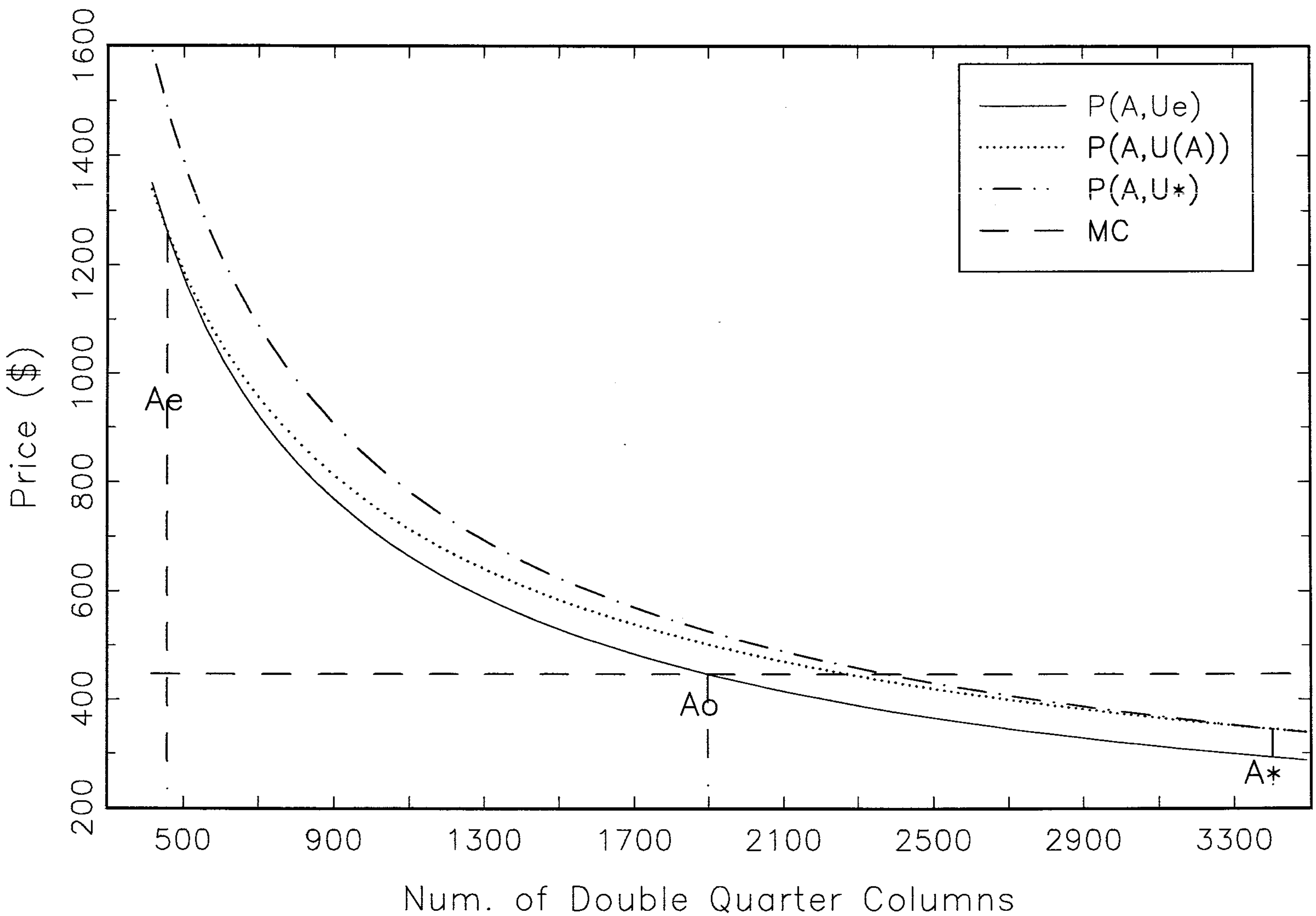
In order to determine the importance of the network effect, I calculate classical and network deadweight loss as discussed in Section 4.4. Figure 2 draws the estimated demand curve for a single directory using parameter estimates from Table 5. The solid line ($P_j(A_j, Ue)$) is the willingness-to-pay curve for the equilibrium choice of advertising Ae. A social planner who did not take advantage of the network effect would choose advertising level Ao. The actual optimal choice is $A*$, where demand ($P_j(A_j, U_j(A_j))$ - the dotted line) is below marginal cost. Choosing $A*$ places the market on the optimal willingness-to-pay curve $P_j(A_j, U*)$ (the dashed-and-dotted line). The space between $P_j(A_j, Ue)$ and $P_j(A_j, U*)$ represents network deadweight loss. Classical deadweight loss is the triangle between $P_j(A_j, Ue)$ and marginal cost MC, to the right of Ae. The results for an average market are computed in Table 3 [27] The ratio of network deadweight loss to classical deadweight loss is 1.26 (note the large standard error: 1.2). Total deadweight loss equals 0.43 of equilibrium consumer surplus (with a standard error of only 0.09). In a market where producers clearly exhibit strong market power, network effects still create a large amount of deadweight loss.

6.1 An Application: Entry

Network effects are at least moderately important in the sense that the positive feedback parameters are statistically significant and network deadweight loss is non-trivial. Are network effects so important that the benefits of monopoly outweigh the benefits of competition? What parameters determine this outcome? This section presents an entry experiment in order to

[27] Standard errors in Tables 3, 7 and 8 are calculated using the delta method.

Figure 2
Deadweight Loss at Estimated Parameters



Network Deadweight Loss		
Summary Variable	Result	Std. Err.
Equilibrium Advt (pages)	418	110
Classical Social Optimum	1,784	506
Social Optimum	3,039	1,511
Equilibrium Surplus ($000's)	25,595	23,054
Class. Soc. Opt. Surplus	30,515	25,439
Soc. Opt. Surplus	36,788	32,535
Classical Deadweight Loss	4,920	2,541
Network Deadweight Loss	6,273	7,725
Ratio of NDWL to CDWL	1.28	1.20
Ratio of Total DWL to Equ Surp.	0.43	0.09

Table 3: Network Deadweight Loss

answer these questions. The results from the entry experiment appear in Table 7. This table presents equilibrium outcomes for different numbers of symmetric competitors that perfectly overlap in an average market. Duopoly firms each choose higher quantity then a monopolist, reflecting the strength of the competitive effect. This result occurs because the estimate of σ is so high. Because there is little differentiation between directories, a book that is slightly larger captures most of the usage market. So two competing directories drive each other to high levels of advertising. Even so, usage at each book drops substantially from the monopoly case. As each publisher enters thereafter, advertising and usage at each directory shrink, so the benefits of the network effect are dissipated. [28] However, total advertising and usage increase, reflecting the benefits of competition.

Table 7 also presents the amount of total surplus generated for each number of competitors. The results show that, ignoring fixed costs, welfare improves in the number of competitors. In this market, network effects are not strong enough to imply that the benefits of monopolization outweigh the benefits of competition. Note that the results of the model could have been

[28] While the non-monotonicity in directory-level advertising is clear in the Table 7, it is difficult to verify in data because the overlapping structure of the distribution areas means that there are almost no clear duopoly or triopoly markets.

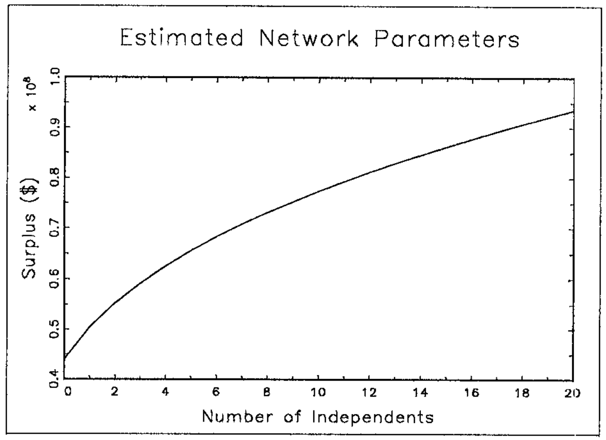
different if the network effect had been estimated to be stronger. Consider raising the network parameter in the advertising demand equation (α_1, the coefficient on usage) and recalculating how surplus changed over the number of competitors. Figure 3 maps mean surplus levels when the parameter is 48% and 55% larger. The figures show that for large network effects, the model implies that welfare decreases in the number of competitors or could even be hump-shaped. Again, these results do not take account of any fixed costs.

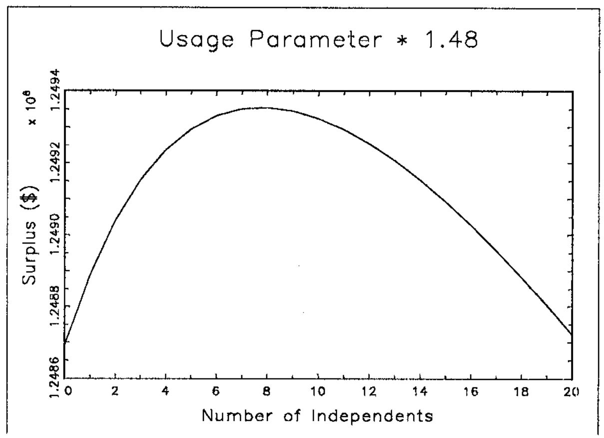
For the actual parameter estimates, surplus increases in the number of competitors. The crucial question for welfare purposes is: how does the increase in surplus due to an entrant compare to the profits of the entrant? Table 8 compares the social benefits of entry to the private benefits captured by the firm. The first column of Table 8 shows that surplus from entrants is considerably higher than their profits. The difference between the increase in total surplus and profits is significantly different then zero for each entrant. This computation is done without considering fixed costs, but almost wherever fixed costs lie, there will be under-entry in equilibrium. This result suggests that current laws that allow entry in the Yellow Pages market should be encouraged.

If we included consumers in the analysis, the result would be even stronger. In this model, seeing that total usage increases in the number of competitors implies that consumer welfare increases in competition. While we cannot convert utils into dollars without observing consumers' response to price, we can use the discrete choice model and parameter estimates to see how much welfare increases in competition. Going from 1 directory to 4 makes consumer welfare from the the Yellow Pages market go up by 22%, and going from 1 to 7 increases welfare from the Yellow Pages market by 35%.

6.2 Equilibrium entry and the optimal number of directories

If we knew the fixed cost of producing a directory, we could push the model further and calculate the number of entrants in equilibrium as well as the optimal number of entrants for a market.





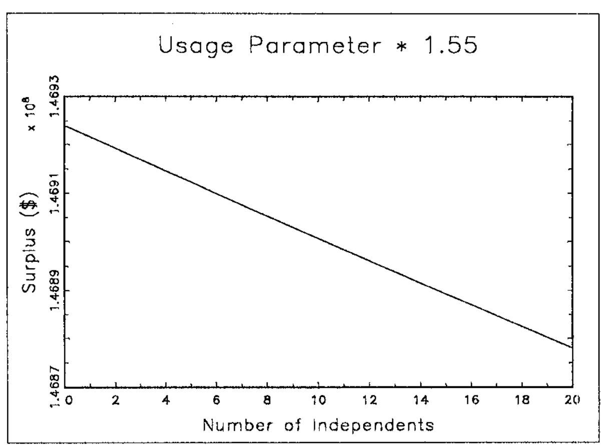


Figure 3: Surplus versus the Number of Competitors.

While obtaining data on the fixed costs of setting up a Yellow Pages directory is difficult, I can obtain a crude estimate by imposing a zero-profit condition on independent directories. The estimate of the fixed cost for independent publishers is $1,004,700 (with a standard error of $765,030). Telephone companies, which are required to distribute White Pages to every phone line, most likely have substantially lower fixed costs for publishing Yellow Pages. The second column of Table 8 computes profits for independents counting this fixed cost. We cannot reject the possibility that only a telephone company directory is profitable in an average-sized market. This prediction seems largely accurate, as we observe entry mostly in larger markets.

In order to calculate the optimal number of entrants, Table 8, Column 3, computes the proportional increase in surplus accounting for the fixed cost of setting up a directory. With 95% confidence, we can say that 4 directories are better than 3 or less. [29] The data have very few markets with 4 or more directories so I take these results to mean that a more competitive market is preferable to the current market structure. As before, measuring consumer utility would mean that the benefits to entry would be even higher.

An interesting question is whether the benefits of entry are driven by increased competition or by increased product choice. The following results show that the effect of competition is the crucial one in this market. First, I removed the competition effect by recalculating the entry experiment assuming that all directories were owned by a single publisher (results not reported). In that case, even having 2 directories is not significantly better than having 1. Second, I mitigated the effects of product differentiation by recalculating the entry experiment for competing directories but with higher values of σ (also not reported). The optimal number of directories decreases as differentiation decreases because fewer consumers switch from the outside option when a new directory enters so advertisers derive less benefit from entry. However, more substitutable directories means that the first two directories drive each other out

[29] Because surplus increases in the number of competitors for any given set of parameters, it is possible for the proportional increase statistic to be statistically different than zero even though 95% confidence intervals around the surplus statistics overlap.

to very high levels of advertising. So reducing differentiation decreases the optimal number of directories but also increases the value of the second entrant, so the optimal number never goes below 2. [30] Therefore, the result that multiple directories are preferable to a single directory would not hold without competing directories but would hold for directories with different levels of differentiation.

A potential problem with these computations is that a steep log-linear demand curve means that there is a large group of retailers who place very high value on each directory. Thus, welfare calculations could depend heavily on the top of the demand curve, where we also have very little data. As an ad hoc investigation of this possibility, I recompute welfare from the outcomes in Table 7, but ignore the "tips" of the demand curve. That is, for welfare calculations, I assume that the demand curve is a horizontal line from the Y-axis over to a kink, and then log-linear according to the estimated demand curve thereafter. Therefore, I calculate welfare as before and then subtract:

$$\int_0^{\bar{A}_j} P(s, U(A_e, A_{-j}))ds - P(\bar{A}_j, U(A_e, A_{-j}))\bar{A}_j$$

where $\bar{A}_j$ is the kink point for directory j and A_e and A_{-j} are equilibrium advertising levels. The last column of table 8 repeats column 3 but with $\bar{A}_j$ set to 30% of equilibrium advertising. For this case, we can no longer say that 4 directories are preferred to 3. However, even for this case, two directories are preferred to one. In fact, to reverse the result that 2 directories are preferred to 1, one would have set the kink at a very high point, well over 40% of equilibrium advertising and well into the range where we have data for most directories. These computations suggest that the result that some competition is preferred to monopoly is not driven by assumptions about the demand curve outside of the range of the data.

[30] A possible criticism of my approach is that the model captures geographic differentiation in ϵ_{ij}. If the model accounted for geographic differentiation explicitly (a difficult endeavor), the estimate of σ might be higher. But this last result suggests that doing so is unlikely to reverse the result that a more competitive market is preferable.

7 Conclusion

This paper examines the welfare trade-off between competition and standardization in a market characterized by network effects. The paper presents a model of the market for Yellow Pages that explicitly captures the relationship between advertising and consumer usage in a directory. The paper estimates the model by extending the techniques of Berry (1994) to the case of overlapping markets with distinct boundaries. The results show that, for a given directory, retailer demand for advertising increases in consumer usage and that consumer demand for directory usage increases in the amount of advertising, implying a network effect. In equilibrium, forgone surplus due to unexploited network effects is high relative to the amount of surplus obtained in equilibrium and also relative to the deadweight loss due to imperfect competition. However, the results show that despite the network effect, a more competitive market structure is preferable to a more concentrated one. Strikingly, the paper finds that multiple entrants improve welfare but are unprofitable in the average market. The results of the paper imply that encouraging competition in the Yellow Pages (as a number of recent policy changes do) improves welfare.

Directory Level Variables		
Name	Description	Instrument Vector
advertising	Number of pages times number of columns times 0.8 if observably smaller, logged. From YPPA library, collected by Boston Consulting Group.	
price	rate for a double-quarter column ad, logged. From *Rate and Data CD-ROM.*	
usage	Number of uses per household per month, logged. From National Yellow Pages Monitor Area Reports.	
Pop. Coverage	Number of people covered by a directory area, logged. From YPPA Industry Characteristics CD-ROM.	U,P,C
telco	Dummy, 1 if directory is associated with phone company. Constructed by observing publisher, books and some company contact.	U,P
GTE, Bell South	Telco Dummies	P,C
Distribution Area	Square mileage of distribution area From Claritas PowerPages CD-ROM	U
County Level Variables - From USA Counties CD-ROM 1996		
urban pop.	% of population in urban setting, 1990	U,P
diff county	% of population in different county 5 years previous, 1990	U,P
diff state	% of population in different state 5 years previous, 1990	U,P
have not moved	% of population in same house as 5 years previous, 1990	U,P
owner occ. house	% of housing that is owner occupied, 1990	U,P
income	per capita income, 1993	U,P
density	population in 1995 divided by square mileage	U,P,C
growth	rate of population increase from 1990 to 1995	U,P
public trans.	% of population regularly using public transportation (1995)	U,P
earnings	per capita earnings	U,P,C

Table 4: Explanation of Variables

8 Appendix

8.1 Equilibrium in the Publisher's Game

This section shows that there exists an equilibrium in pure strategies in the publishers' game. I show that each publisher's objective function is concave in its choice variable and then I appeal to Theorem 1.2 in Fudenberg and Tirole (1991), originally proven by Debreu, Glicksberg and Fan. Publisher j, $j = 1, ..., J$, chooses its quantity of advertising A_j to maximize profits: $P_j(A_j, U_j(A_1, ..., A_J))A_j - C_j(A_j)$. For this sub-section, I assume $P_j = A_j^{\gamma} U_j^{\alpha_1} X_j$. The variable X_j captures directory level characteristics. To parameterize $U(.)$, I assume a logit model holds in each sub-market, where mean utility to directory j is $\delta_j = \alpha_2 \ln(A_j) + \ln(X_j)$. Sub-markets (described in Section 4.2) are regions with a uniform set of directories. Let $K(j)$

be the set of sub-markets that are covered by directory j and let $C(k)$ be the set of directories that serve sub-market k. Let ψ_k be the share of j's market represented by sub-market k (the j will be obvious in context). Therefore:

$$s_{jk} = \frac{A_j^{\alpha_2} X_j}{1 + \sum_{i \in C(k)} A_i^{\alpha_2} X_i} \qquad s_j = \sum_{k \in K(j)} \psi_k s_{jk}$$

Usage is defined by $U_j = M s_j$, where M is the size of the market. The parameterized version of the publisher first-order (Equation 5) is:

$$P_j \left(1 + \gamma + \frac{\alpha_1 \alpha_2}{s_j} \sum_{k \in K(j)} \psi_k s_{jk}(1 - s_{jk}) \right) = MC_j.$$

The term in large parentheses represents the price-cost markup. Allow Γ_j to equal the markup, so $P_j \Gamma_j = MC_j$. Concavity requires that the second derivative is negative:

$$\left(\frac{\partial P_j}{\partial A_j} + \frac{\partial P_j}{\partial U_j} \frac{\partial U_j}{\partial A_j} \right) \Gamma_j + P_j \alpha_1 \alpha_2 \frac{\partial}{\partial A_j} \frac{\sum_{k \in K(j)} \psi_k s_{jk}(1 - s_{jk})}{s_j} < 0. \tag{7}$$

The term in the first set of parentheses is the slope of the demand curve and is assumed to be negative. The price-cost margin Γ_j is positive for any reasonable parameter values. In the next term, P_j, α_1 and α_2 are each positive. Unfortunately, the term $\frac{\partial}{\partial A_j} \frac{\sum_{k \in K(j)} \psi_k s_{jk}(1-s_{jk})}{s_j}$ is difficult to sign. In the case of perfect overlap, that term becomes $\frac{\partial}{\partial A_j}(1 - s_j)$ which is obviously negative, so there is a solution in pure strategies. In the case without perfect overlap, the term is:

$$\frac{\partial}{\partial A_j} \frac{\sum_{k \in K(j)} \psi_k s_{jk}(1 - s_{jk})}{s_j} = -\frac{[\sum_{k \in K(j)} \psi_k s_{jk}(1 - s_{jk})]^2}{s_j} + \sum_{k \in K(j)} (1 - 2s_{jk})(1 - s_{jk}) s_{jk}.$$

The first term on the right-hand side is always negative but the second term might be positive. I must assume that α_1 and α_2 are not "too large" to guarantee an equilibrium in pure strategies. Condition 7 easily holds at the parameters estimated in this paper.

8.2 Computational Details of the Usage Equation

In order to write down the correct relationship between s_j and δ_j, let $K(j)$ be the set of sub-markets in the region covered by directory j, let ψ_{jk} be the portion of directory j's market in sub-market k, let s_{0k} be the share of reference to the outside option in sub-market k and let $s_{j|YPk}$ be the share to j in the Yellow Pages group. In this case, we have:

$$s_j = e^{\delta_j} \sum_{k \in K(j)} \psi_{jk} s_{0k} s_{j|YPk}^{\sigma}. \tag{8}$$

I do not observe $s_{j|YPk}$ or s_{0k}. However, the assumptions on functional form that are a part of the nested logit model imply:

$$s_{j|YPk} = \frac{e^{\delta_j/(1-\sigma)}}{\sum_{l \in C(k)} e^{\delta_l/(1-\sigma)}} \qquad s_{0k} = \frac{1}{1 + (\sum_{l \in C(k)} e^{\delta_l/1-\sigma})^{\sigma}}. \tag{9}$$

$C(k)$ is the set of directories which compete in sub-market k. The equations in 9 come from Cardell (1997) and Berry (1994) We can compute s_j as a function of δ by plugging 9 into 8. I cannot solve for δ explicitly in this case. I use a fixed-point algorithm to solve this problem. Let s be the vector of observed market shares and let $s(\delta)$ by the vector of predicted market shares defined by Equations 8 and 9. Define the function $g(.)$ by:

$$g(\delta) = \delta + s - s(\delta) \tag{10}$$

Below, I show that Equation 10 is a contraction mapping for any σ below a certain cut-off, ensuring that a unique δ exists and that it can be found. I check that σ is less than the cut-off before each implementation of the fixed point algorithm.

In order to increase identification power over σ, I use the fact that $s_{j|YP} = s_{j|YPk}$ and $s_0 = s_{0k}$ at markets with only one sub-market. Thus, I can combine information from Equation 4 and Equation 8. I use the following approach in order to construct ξ: Start with a given σ, α_1 and β_u. Use σ and Equations 8, 9 and 10 to obtain δ via the fixed point algorithm. Let $\kappa_j = 1$ if directory j has one sub-market and at least one competitor, and $\kappa_j = 0$ otherwise. (Note that markets with no competitors confer no information about σ). Define ξ_j by:

$$\xi_j = \begin{cases} \ln(s_j) - \ln(s_0) - \sigma \ln(s_{j|YP}) - \alpha_1 \ln(A_j) - X_j\beta_u & \text{if } \kappa_j = 1 \\ \delta_j - \alpha_1 \ln(A_j) - X_j\beta_u & \text{if } \kappa_j = 0 \end{cases}$$

In this data, 121 directories out of 428 have a uniform set of directories offered across their entire market. Of those, 65 do not overlap with any other directories and 56 are completely overlapped by all competitors.

Now I turn to establishing that Equation 8 has a unique fixed point. I show that the equation $g : \Re^J \to \Re^J$ defined as $g(\delta) = \delta + s - s(\delta)$ is a contraction by showing that the function satisfies the conditions stated in the theorem in Appendix I of Berry, Levinsohn and Pakes (1995). The important conditions to show are that $g(\delta)$ is continuous in δ, that $\partial g_j(\delta)/\partial\delta_i \geq 0$ for all i and j, and that $\sum_{i=1}^{J} \partial g_j(\delta)/\partial\delta_i < 1$. The function is continuous by construction. I show that $\partial g_j(\delta)/\partial\delta_j \geq 0$ last as this part of the proof requires conditions on σ.

First, I establish that $\partial g_j(\delta)/\partial\delta_i \geq 0$ for $i \neq j$.

$$\frac{\partial g_j(\delta)}{\partial\delta_i} = -e^{\delta_j} \sum_{k\in K(j)} \psi_{jk} \frac{\partial}{\partial\delta_i} s_{0k} s^{\sigma}_{j|YPk}.$$

Following from Berry (1994), it is easy to show that $\partial s_{j|YPk}/\partial\delta_i = -s_{j|YPk}s_{i|YPk}/(1-\sigma)$ and that $\partial s_{0k}/\partial\delta_i = -s_{ik}s_{0k}$. Plugging in, we have that:

$$\frac{\partial g_j(\delta)}{\partial\delta_i} = e^{\delta_j} \sum_{k\in K(j)} \psi_{jk} s_{0k} s^{\sigma}_{j|YPk} \left(\frac{\sigma}{1-\sigma} s_{i|YPk} + s_{ik} \right) > 0.$$

In order to show that $\sum_{i=1}^{J} \partial g_j(\delta)/\partial\delta_i < 1$, note that $\partial s_{j|YPk}/\partial\delta_i = (s_{j|YPk} - s^2_{j|YPk})/(1-\sigma)$. Therefore:

$$\frac{\partial g_j(\delta)}{\partial\delta_j} = 1 - s(\delta) + e^{\delta_j} \sum_{k\in K(j)} \psi_{jk} s_{0k} s^{\sigma}_{j|YPk} \left(\frac{\sigma}{1-\sigma}(s_{j|YPk} - 1) + s_{jk} \right)$$

$$\sum_{i=1}^{J} \frac{\partial g_j(\delta)}{\partial \delta_i} = 1 - s(\delta) + e^{\delta_j} \sum_{k \in K(j)} \psi_{jk} s_{0k} s^{\sigma}_{j|YPk}(1 - s_{0k}) =$$

$$1 - e^{\delta_j} \sum_{k \in K(j)} \psi_{jk} s^2_{0k} s_{j|YPk} < 1.$$

Now I establish conditions that insure that $\partial g_j(\delta)/\partial \delta_j > 0$. It is equivalent to show that $\partial s_j(\delta)/\partial \delta_j < 1$. A sufficient condition is that $\partial s_{jk}(\delta)/\partial \delta_j < 1$ for all sub-regions k. Suppressing the subscript k, I show that:

$$\frac{\partial s_j}{\partial \delta_j} = \frac{\partial s_{j|YP} s_{YP}}{\partial \delta_j} = s_{j|YP} \frac{\partial s_{YP}}{\partial \delta_j} + s_{YP} \frac{\partial s_{j|YP}}{\partial \delta_j} < 1.$$

Following Berry (1994), we have that $\partial s_{j|YP}/\partial \delta_j = s_{j|YP}(1 - s_{j|YP})/(1 - \sigma)$. Also, we have that $s_{YP} = \exp(x)/1 + \exp(x)$ where $x = (1 - \sigma) \ln \sum_j \exp(\delta_j/(1 - \sigma))$. The derivative reduces to $\partial s_{YP}/\partial \delta_j = s_{YP}(1 - s_{YP}) s_{j|YP}$. Plugging into $\partial s_j/\partial \delta_j$ and solving for σ shows that we have sufficient conditions for a contraction whenever:

$$\sigma < \frac{1 - s_j(1 - s_j)}{1 - s_j(s_{j|YP} - s_j)}. \tag{11}$$

It is easiest to study this condition by converting $s_j = s_{j|YP} s_{YP}$, as the latter two terms can be moved independently of each other. The upper bound on σ decreases in s_{YP} and is convex in $s_{j|YP}$, reaching a minimum at $s_{j|YP} = 0.5$. The bound is always greater than 0.75 and less than 1. There is actually some intuition to the result that σ must be less than one. We are trying to show that $\partial s_j/\partial \delta_j$ is less than one. When σ is high, all of the randomness in utility is placed at the group level, which means that within-group choices are based almost entirely on mean utility. When mean utility (δ_j) moves slightly, it generates a big response and s_j rises too quickly. Of course, if few people choose the group, σ can be higher and s_j still will rise at a reasonable rate. And as is typical in logit models, the within-group derivative is highest when the within-group market share is close to 0.5.

I do not impose the bound in estimation. Instead, I check for σ to satisfy (11) at iteration of my optimization routine, before searching for a fixed point. Note that I always start my estimation procedure with a guess of σ that is below 0.75, ensuring that a fixed point exists at my starting values. If the true σ is greater than the bound, I expect my estimate of σ to rise. When the estimate of σ crosses the bound, my estimation routine stops because estimates of δ might be meaningless at that point.

9 Bibliography

ANDERSON, S. and COATE, S. (2001), "Market Provision of Public Goods: The Case of Broadcasting", mimeo, Cornell University.

ARMSTRONG, M. (2002), "Competition in Two-Sided Markets", mimeo, Oxford University.

BERRY, S. (1994), "Estimating Discrete Choice Models of Product Differentiation", *Rand Journal of Economics*, 25, 242-262.

BERRY, S., Levinsohn, J. and Pakes, A. (1995),"Automobile Prices in Market Equilibrium", *Econometrica*, 63, 841-890.

BERRY, S. and WALDFOGEL, J. (1999), "Free Entry and Social Inefficiency in Radio Broadcasting.", *Rand Journal of Economics*, 30, 397-420.

CARDELL, N.S. (1997), "Variance Components Structures for the Extreme Value of Logistic Distributions with Application to Models of Heterogeneity", *Econometric Theory*, 113, 185-213.

CHOU, C. and SHY, O. (1990) "Network Effects without Network Externalities", *International Journal of Industrial Organization*, 8, 259-70.

CHURCH, J. and GANDAL, N. (1992), "Network Effects, Software Provision, and Standardization", *Journal of Industrial Economics*, 40, 85-103.

ECONOMIDES, N. (1996), "The Economics of Networks", *International Journal of Industrial Organization*, 14, 673-99.

ECONOMIDES, N. and FLYER, F. (1997), "Compatibility and Market Structure for Network Goods", (manuscript, Stern School of Business at New York University).

ECONOMIDES, N. and HIMMELBERG, C. (1995), "Critical Mass and Network Size with Application to the US FAX Market", (manuscript, Stern School of Business at New York University).

ELLIOTT, S. (1998), "Advertising", *New York Times*, April 17, C7.

Feist Publications, Inc., v. Rural Telephone Service Company, Inc. *Supreme Court Reporter*, 111 S.Ct. 1282(1991).

FUDENBERG, D. and TIROLE, J. (1991), *Game Theory* (Cambridge: MIT Press).

GANDAL, N. (1994), "Hedonic Price Indexes for Spreadsheets and an Empirical Test of Network Externalities", *Rand Journal of Economics*, 25, 161-170.

GANDAL, N., GREENSTEIN, S.M. and SALANT, D. (1999) "Adoptions and Orphans in the Early Microcomputer Market", *Journal of Industrial Economics*, 47, 87-105.

GANDAL, N., KENDE, M. and ROB, R. (2000) "The Dynamics of Technological Adoption in

Hardware/Software Systems: The Case of Compact Disk Players", *Rand Journal of Economics*, 31, 43-62.

GOOLSBEE, A. and KLENOW, P. (1999), "Evidence on Learning and Network Externalities in the Diffusion of Home Computers", (NBER Working Paper 7329, September).

GREENSTEIN, S. M. (1993), "Did An Installed Base Give an Incumbent any (Measurable) Advantage in Federal Computer Procurement?" *Rand Journal of Economics*, 24, 19-39.

HANSEN, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

KATZ, M.L. and SHAPIRO, C. (1985), "Network Externalities, Competition, and Compatibility", *American Economic Review*, 75, 424-40.

KATZ, M.L. and SHAPIRO, C. (1994), "Systems Competition and Network Effects", *Journal of Economic Perspectives*, 8, 93-115.

LABAND, D.N. (1986), "Advertising as Information: An Empirical Note", *Review of Economics and Statistics*, 68, 517-521.

LIEBOWITZ, S.J. and Margolis, S.E. (1994), "Network Externality: An Uncommon Tragedy", *Journal of Economic Perspectives*, 8, 133-150.

MIXON, F.G. (1995), "Advertising as Information: Further Evidence", *Southern Economic Journal*, 61, 1213-1218.

NARUC, *Compilation of Utility Regulatory Policy 1994-1995*, National Association of Regulatory Utility Commissioners. pp. 376.

OHASHI, H. (2000), "Network Externailities and Consumer Welfare: Home Video Cassett Recorders in the U.S., 1978-1986", (manuscript, University of British Columbia).

PARK, S. (2000), "Quantitative Analysis of Network Externalities in Competing Technologies: The VCR Case", (manuscript, SUNY at Stony Brook).

ROCHET, J.-C. and TIROLE, J. (2001), "Platform Competition in Two-Sided Markets," mimeo, University of Toulouse.

ROSSE, J.N. (1970), "Estimating Cost Function Parameters without Using Cost Data: Illustrated Methodology", *Econometrica*, 38(2), 256-275.

RUST, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 55, 999-1033.

SALONER, G. and SHEPARD, A. (1995), "Adoption of Technologies With Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines", *Rand Journal of Economics*, 26, 479-501.

STEGEMAN, M. (2002), "Competitive Provision of Non-rival and Pure Public Goods Through Advertising", mimeo, Virginia Polytechnic Institute.

WHITE, E.D. and SHEEHAN, M.F. (1992), "Monopoly, The Holding Company, and Asset Stripping: The Case of Yellow Pages", *Journal of Economic Issues*, 26, 159-182.

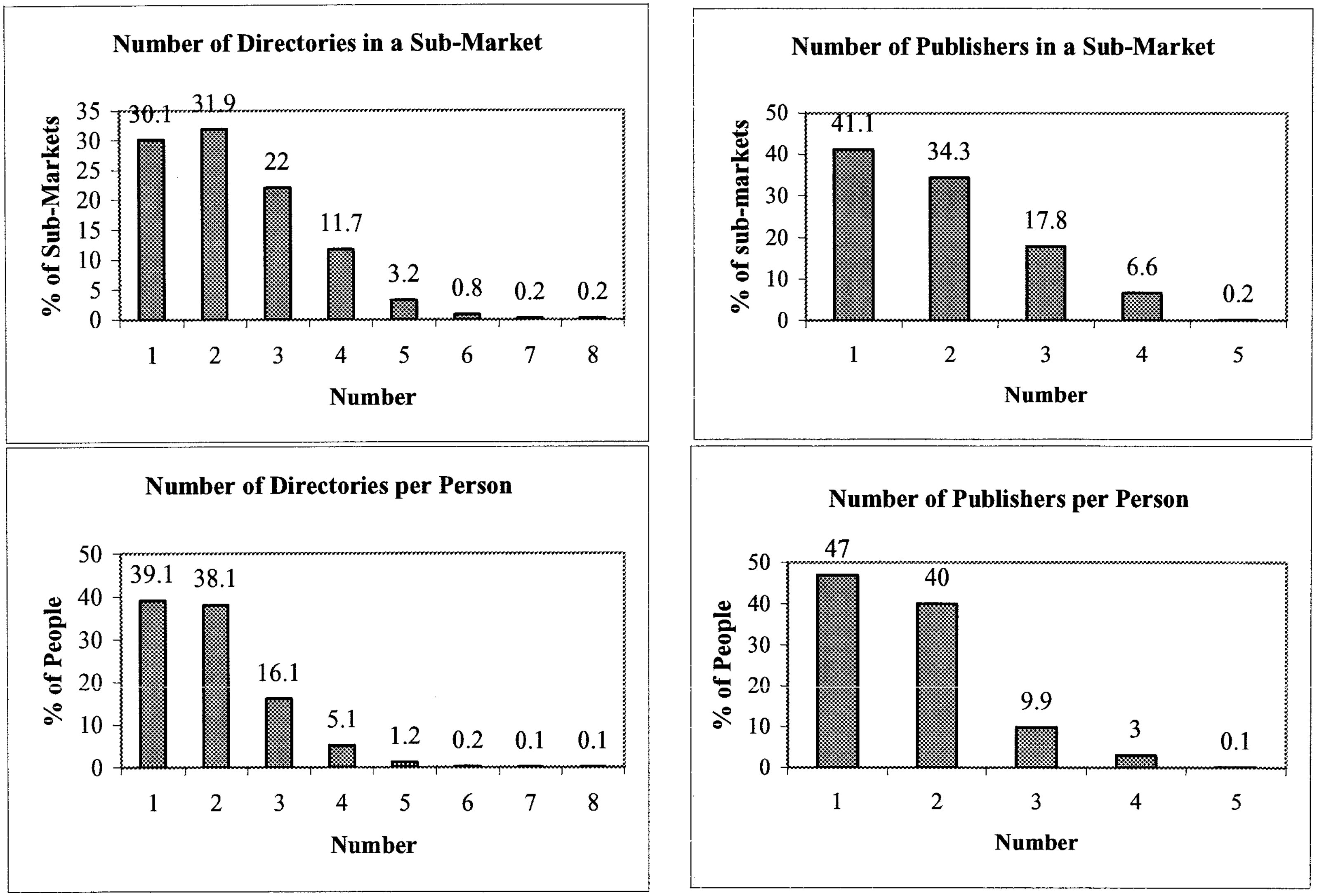

Number of Directories in a Sub-Market
% of Sub-Markets
35
30
25
20
15
10
5
0
30.1
31.9
22
11.7
3.2
0.8
0.2
0.2
1
2
3
4
5
6
7
8
Number
Number of Publishers in a Sub-Market
% of sub-markets
50
40
30
20
10
0
41.1
34.3
17.8
6.6
0.2
1
2
3
4
5
Number
Number of Directories per Person
% of People
50
40
30
20
10
0
39.1
38.1
16.1
5.1
1.2
0.2
0.1
0.1
1
2
3
4
5
6
7
8
Number
Number of Publishers per Person
% of People
50
40
30
20
10
0
47
40
9.9
3
0.1
1
2
3
4
5
Number

Figure 4

Table 4: "First Stage" Regressions

These are OLS regressions of endogenous variables on their respective instruments.
Note that in GMM, there is no explicit "first stage".

Usage Equation Instruments

Dependent Variable:	Advertising		Within Group Share (Sj|YP)	
Population Coverage	0.645	(0.041)	0.547	(0.065)
Distribution Area	0.078	(0.026)	0.370	(0.039)
constant	2.164	(0.72)	-5.238	(1.577)
% urban population	0.005	(0.00)	0.001	(0.006)
% lived in diff county	0.045	(0.01)	0.002	(0.019)
% lived in diff state	0.035	(0.01)	-0.007	(0.024)
% own house	0.020	(0.01)	0.021	(0.015)
% grad hi school	-0.025	(0.01)	-0.047	(0.016)
% grad college	-0.012	(0.01)	-0.099	(0.024)
per cap income	0.050	(0.02)	0.171	(0.038)
telco book	0.681	(0.06)	1.666	(0.093)
# building permits	-0.059	(0.051)	-0.203	(0.089)
county growth rate	-0.010	(0.012)	0.010	(0.030)
% take public trans.	0.008	(0.026)	0.149	(0.040)
% have not moved	-0.008	(0.012)	-0.033	(0.027)
pop. density	-2.80E-05	(3.33E-05)	-2.54E-04	(4.65E-05)
F-stat for bold variables	20.94		13.10	
R-Squared	0.60		0.76	

Price Equation Instruments

Dependent Variable:	Advertising		Usage	
Earnings per worker	-0.020	(0.009)	-0.019	(0.014)
Bell South	0.093	(0.143)	0.338	(0.218)
GTE	0.041	(0.087)	-0.163	(0.133)
% have not moved	0.011	(0.007)	0.040	(0.010)
% lived in diff county	0.057	(0.009)	0.087	(0.014)
% lived in diff state	0.048	(0.011)	0.065	(0.016)
constant	1.878	(0.574)	-2.463	(0.874)
% urban population	0.003	(0.002)	-0.005	(0.004)
% grad hi school	-0.017	(0.008)	-0.037	(0.013)
% grad college	-0.022	(0.011)	-0.039	(0.017)
per cap income	0.060	(0.019)	0.078	(0.030)
telco book	0.683	(0.062)	1.547	(0.095)
population coverage	0.699	(0.039)	0.342	(0.059)
establishments per cap	0.141	(0.100)	0.110	(0.152)
population density	-2.94E-05	(2.68E-05)	-7.80E-05	(4.08E-05)
F-stat for bold variables	3.56		3.33	
R-Squared	0.59		0.54	

Bold variables are "excluded".
Standard errors are in parenthesis.

Table 5: Results from the Generalized Method of Moments

	Dependent Variable	Mrg Efct	coef.	std error
PRICE OF	quantity of advertising (γ)		-0.729	0.193
ADVERTISING	usage (α1)		0.564	0.131
	constant		6.735	0.715
	population coverage		0.741	0.100
	telco book		0.013	0.138
	establishments per cap.		0.026	0.064
	% urban population		-2.02E-04	2.25E-03
	% grad college		0.009	0.009
	% grad hi school		0.009	0.006
	per cap. Income		0.014	0.014
	pop. density		2.68E-05	1.44E-05
USAGE	advertising (α2)	0.001	0.154	0.073
(ln(sj/s0))	constant		-2.964	0.878
	% urban population	0.024	0.003	0.003
	% lived in diff county	0.181	0.022	0.008
	% lived in diff state	0.170	0.020	0.008
	% have not moved	0.010	0.009	0.010
	% own house	-0.104	-6.05E-03	0.007
	% grad hi school	-0.104	-0.012	0.007
	% grad college	-0.133	-0.016	0.008
	per cap income	0.065	0.008	0.013
	telco book	2.536	0.304	0.093
	# of house-building permits	-0.598	-0.072	0.032
	county pop. growth rate	0.129	0.015	0.010
	% take public trans.	0.053	0.006	0.015
	pop. density	-4.20E-04	-5.03E-05	1.88E-05
MARGINAL	constant		3.228	0.677
REVENUE	population coverage		0.437	0.116
	earnings per worker		0.003	0.014
	pop. density		9.65E-05	4.03E-05
	Bell South		-0.631	0.529
	GTE		0.612	0.129
CORRELATION (sigma)			0.803	0.079
p-value of exogeneity test, d.f			0.88	4
Number of Observations				428

Table 6: Robustness Results

	No Instrumenting		Market Size (M) = 35		w/ Competition Vars.	
	coef.	std error	coef.	std error	coef.	std error
PRICE EQUATION						
advertising	-0.004	0.035	-0.727	0.192	-0.757	0.501
usage	0.116	0.020	0.562	0.130	0.762	0.521
constant	4.435	0.301	6.730	0.712	6.762	0.976
population coverage	0.392	0.035	0.741	0.100	0.663	0.194
telco book	0.216	0.047	0.017	0.137	-0.362	0.402
establishments per cap.	0.063	0.045	0.025	0.064	0.061	0.094
% urban population	-0.004	0.001	-1.84E-04	2.24E-03	0.001	0.004
% grad college	-0.004	0.005	0.009	0.009	0.004	0.011
% grad hi school	0.002	0.004	0.009	0.006	0.011	0.011
per cap. Income	0.025	0.010	0.014	0.014	0.024	0.020
pop. density	2.33E-05	1.44E-05	2.63E-05	1.44E-05	3.27E-05	1.79E-05
ads at competing books					0.062	0.080
usage at competing books					-0.103	0.158
USAGE EQUATION						
advertising	0.230	0.034	0.158	0.082	0.146	0.074
constant	-2.419	0.639	-1.800	1.007	-2.987	0.883
% urban population	0.003	0.003	0.004	0.003	0.003	0.003
% lived in diff county	0.025	0.008	0.026	0.009	0.022	0.008
% lived in diff state	0.020	0.010	0.023	0.010	0.021	0.009
% have not moved	0.012	0.010	0.008	0.012	0.010	0.011
% own house	-0.009	0.007	-0.006	0.008	-0.006	0.007
% grad hi school	-0.011	0.009	-0.016	0.009	-0.011	0.007
% grad college	-0.015	0.009	-0.016	0.010	-0.015	0.008
per cap income	3.30E-04	1.43E-02	0.006	0.016	0.004	0.014
telco book	0.260	0.059	0.329	0.107	0.304	0.094
# of house-building permits	-0.094	0.038	-0.082	0.037	-0.079	0.032
county pop. growth rate	0.019	0.010	0.015	0.011	0.015	0.010
% take public trans.	0.006	0.017	0.010	0.018	0.005	0.015
pop. density	-5.89E-05	2.16E-05	-6.35E-05	2.26E-05	-4.87E-05	1.92E-05
COST EQUATION						
constant	4.333	0.367	3.154	0.936	3.140	0.890
population coverage	0.521	0.057	0.438	0.113	0.445	0.121
earnings per worker	-0.004	0.011	0.009	0.026	0.006	0.015
pop. density	1.41E-05	2.06E-05	8.38E-05	6.55E-05	8.75E-05	4.71E-05
Bell South	0.587	0.083	-0.424	1.272	-0.555	0.672
GTE	0.121	0.086	0.505	0.203	0.632	0.200
CORRELATION (sigma)	0.8	Fixed	0.0796	0.091	0.807	0.079
p-value of exogeneity test, d.f.	Just Identified		0.9125	4	0.91	2
Number of Observations		428		428		428

Surplus Increase minus Profits (%) is *(incsurp(k,k-1)-prof(k))/incsurp(k,k-1)*

Table 7
Equilibrium for Different Numbers of Competitors

# of competitors	advertising (pages)		refs./HH/mth.		price ($) (DQC ad)		profits ($)*		Total Surplus*	
1	613	(578)	4.10	(0.69)	2,136	(1,207)	5.16	(1.60)	26.61	(19.67)
2	707	(606)	2.38	(0.38)	1,416	(794)	2.85	(1.00)	38.50	(29.45)
3	624	(533)	1.68	(0.28)	1,273	(736)	1.97	(0.79)	45.00	(35.06)
4	549	(470)	1.30	(0.22)	1,212	(712)	1.53	(0.68)	49.74	(39.39)
5	490	(420)	1.07	(0.19)	1,178	(699)	1.26	(0.60)	53.55	(43.01)
6	443	(381)	0.91	(0.16)	1,156	(690)	1.08	(0.55)	56.79	(46.18)
7	405	(349)	0.79	(0.15)	1,141	(684)	0.95	(0.50)	59.62	(49.02)

*Profits and surplus are in millions.
*Profits and surplus are computed assuming there are no fixed costs of production.
Standard Errors are in parenthesis.

Table 8
Private Returns vs. Social Returns

# of competitors	Surplus Increase minus Profits (%) (no fixed costs)		Profits (incl. fixed costs)		Surplus Increase (%) (incl. fixed costs)		Adjusted Surplus Increase (%) (incl. fixed costs)	
2	0.76	(0.17)	1.80	(1.15)	0.42	(0.11)	0.26	(0.11)
3	0.70	(0.22)	0.92	(0.98)	0.15	(0.06)	0.07	(0.08)
4	0.68	(0.25)	0.48	(0.90)	0.09	(0.04)	0.03	(0.07)
5	0.67	(0.26)	0.21	(0.85)	0.06	(0.03)	0.01	(0.06)
6	0.67	(0.27)	0.03	(0.82)	0.05	(0.03)	0.00	(0.06)
7	0.66	(0.27)	-0.10	(0.80)	0.04	(0.03)	-0.01	(0.06)

Surplus Increase minus Profits (%) is *(incsurp(k,k-1)-prof(k))/incsurp(k,k-1)*
Surplus Increase (%) is *(incsurp(k,k-1))/surp(k-1))*
where *surp(k)* equals surplus generated by *k* competitors
incsurp(k,k-1)=surp(k)-surp(k-1)
prof(k) is profit when there are *k* competitors
Adjusted Surplus is computed ignoring the upper tip of the demand curve.
Standard Errors are in parenthesis.